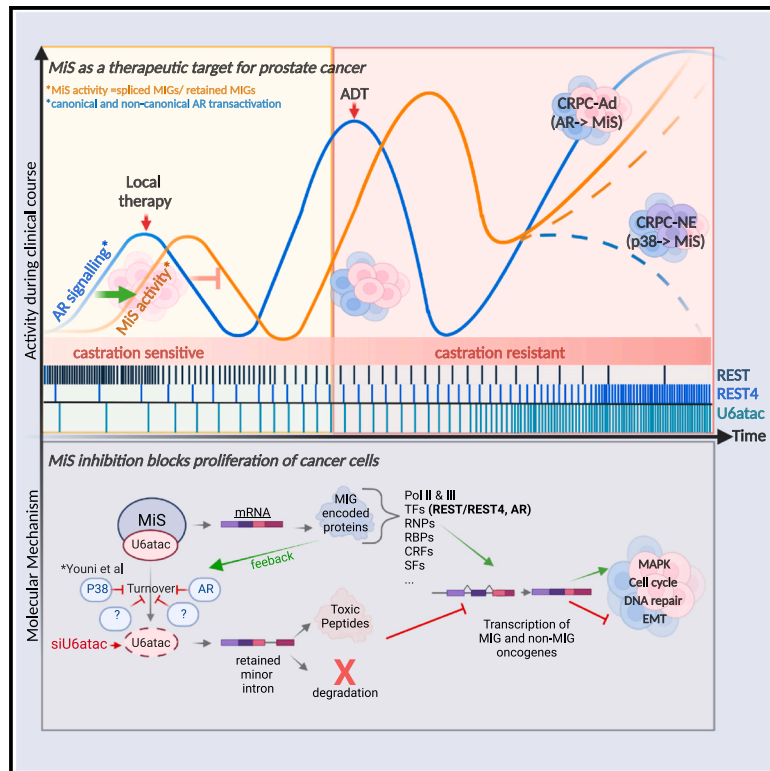# Minor intron splicing is critical for survival of lethal prostate cancer

Graphical abstract

Authors

Anke Augspach, Kyle D. Drake,
Luca Roma, ..., Salvatore Piscuoglio,
Rahul N. Kanadia, Mark A. Rubin

Correspondence

rahul.kanadia@uconn.edu (R.N.K.),
mark.rubin@unibe.ch (M.A.R.)

In brief

Augspach et al. discover a significant vulnerability in therapy-resistant prostate cancer. They show that RNAi-based inhibition of the minor spliceosome blocked the cell cycle and reduced survival in cancer cells, whereas it did not affect non-cancer cells. This approach was more effective than current state-of-the-art therapies.

Highlights

- Increased minor spliceosome activity is associated with prostate cancer progression

- Minor intron-containing genes are overrepresented as direct interactors with oncogenes

- Minor splicing alters splice patterns of important prostate cancer drivers

- Downregulating the minor spliceosome creates a prostate cancer vulnerability

CellPress

Molecular Cell

Article

Minor intron splicing is critical for survival of lethal prostate cancer

Anke Augspach,[1] Kyle D. Drake,[2] Luca Roma,[3] Ellen Qian,[4,5] Se Ri Lee,[4,5] Declan Clarke,[6,7] Sushant Kumar,[6,7,22,23] Muriel Jaquet,[1] John Gallon,[3] Marco Bolis,[8,9] Joanna Triscott,[1] José A. Galván,[10] Yu Chen,[11,12,13] George N. Thalmann,[1,14] Marianna Kruithof-de Julio,[1,14,15] Jean-Philippe P. Theurillat,[8,16] Stefan Wuchty,[17,18,19] Mark Gerstein,[4,6,7] Salvatore Piscuoglio,[3,20] Rahul N. Kanadia,[2,21,24,*] and Mark A. Rubin[1,15,24,25,*]

[1]Department for BioMedical Research, University of Bern, 3008 Bern, Switzerland
[2]Physiology and Neurobiology Department, University of Connecticut, Storrs, CT 06269, USA
[3]Institute of Pathology and Medical Genetics, University Hospital Basel, 4056 Basel, Switzerland
[4]Department of Computer Science, Yale University, New Haven, CT 06520, USA
[5]Yale College, New Haven, CT 06520, USA
[6]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA
[7]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA
[8]Institute of Oncology Research, 6500 Bellinzona, Switzerland
[9]Computational Oncology Unit, Department of Oncology, Istituto di Ricerche Farmacologiche "Mario Negri" IRCCS, 20156 Milano, Italy
[10]Institute of Pathology, University of Bern, Bern 3008, Switzerland
[11]Human Oncology and Pathogenesis Program (HOPP), Memorial Sloan Kettering, New York, NY 10065, USA
[12]Department of Medicine, Memorial Sloan Kettering, New York, NY 10065, USA
[13]Department of Medicine, Weill Cornell Medicine, New York, NY 10065, USA
[14]Department of Urology, Inselspital, Bern University Hospital, 3008 Bern, Switzerland
[15]Bern Center for Precision Medicine, University of Bern and Inselspital, 3008 Bern, Switzerland
[16]Faculty of Biomedical Sciences, Università della Svizzera italiana, 6900 Lugano, Switzerland
[17]Department of Computer Science, University of Miami, Coral Gables, FL 33146, USA
[18]Sylvester Comprehensive Cancer Center, University of Miami, Coral Gables, FL 33136, USA
[19]Department of Biology, University of Miami, Coral Gables, FL 33146, USA
[20]Department of Biomedicine, University of Basel, 4031 Basel, Switzerland
[21]Institute for Systems Genomics, University of Connecticut, Storrs, CT, USA
[22]Present address: Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 1L7, Canada
[23]Present address: Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 0A3, Canada
[24]Senior author
[25]Lead contact
*Correspondence: rahul.kanadia@uconn.edu (R.N.K.), mark.rubin@unibe.ch (M.A.R.)
https://doi.org/10.1016/j.molcel.2023.05.017

SUMMARY

The evolutionarily conserved minor spliceosome (MiS) is required for protein expression of ∼714 minor intron-containing genes (MIGs) crucial for cell-cycle regulation, DNA repair, and MAP-kinase signaling. We explored the role of MIGs and MiS in cancer, taking prostate cancer (PCa) as an exemplar. Both androgen receptor signaling and elevated levels of U6atac, a MiS small nuclear RNA, regulate MiS activity, which is highest in advanced metastatic PCa. siU6atac-mediated MiS inhibition in PCa *in vitro* model systems resulted in aberrant minor intron splicing leading to cell-cycle G1 arrest. Small interfering RNA knocking down U6atac was ∼50% more efficient in lowering tumor burden in models of advanced therapy-resistant PCa compared with standard antiandrogen therapy. In lethal PCa, siU6atac disrupted the splicing of a crucial lineage dependency factor, the RE1-silencing factor (REST). Taken together, we have nominated MiS as a vulnerability for lethal PCa and potentially other cancers.

INTRODUCTION

Splicing results in the removal of introns and the ligation of exons to produce mRNA encoding a full-length protein. A majority (>99.5%) of the introns share consensus sequences at the 5′ splice site (SS), the branch point sequence (BPS), and 3′ SS, which are recognized by small nuclear RNAs (snRNAs) plus proteins of the major spliceosome (U1, U2, U4, U5, and U6 snRNAs). However, <0.5% set of introns possess divergent 5′ SS, BPS, and 3′ SS consensus sequences requiring an alternate spliceosome (U11, U12, U4atac, U5, and U6atac snRNAs) called the minor spliceosome (MiS). Despite the small number (∼714) of minor intron-containing genes

(MIGs), they are highly enriched in critical biological processes, including cell proliferation and differentiation, which is reflected in many developmental diseases, such as microcephalic osteodysplastic primordial dwarfism type 1 (MOPD1), Rofiman syndrome, early onset cerebellar ataxia (EOCA), isolated growth hormone defficiency (IGHD), and others.[1–6] As minor introns are also present in many critical oncogenes, a thorough understanding of the interplay between minor intron splicing and cancer is warranted.[7] There is a reliable association of MiS components with an increased risk of scleroderma (U11/U12-65K protein), acute myeloid lukemia (AML) (U11-59K protein), and familial prostate cancer (PCa) (U11 snRNA).[8–10] Given the above, we posit that the MiS and MIG expression play vital roles in cancer progression, a hypothesis we explore here using PCa as an exemplar cancer.

Although androgen deprivation therapy (ADT) is often used to treat advanced PCa, some patients develop ADT-resistant PCa, which is treated with androgen receptor signaling inhibitors (ARSi), such as enzalutamide and abiraterone. Unfortunately, intrinsic or acquired resistance to ARSi in the form of castration-resistant prostate cancer (CRPC) eventually develops in this ultimately lethal disease. In CRPC, ARSi-resistance can occur via intra-tumoral heterogeneity driven by re-activation of the AR axis[11,12] or transdifferentiation (also known as lineage plasticity) from adenocarcinoma (CRPC-adeno) to neuroendocrine PCa (CRPC-NE), which is highly lethal.[13]

Despite sharing the genomic landscape, CRPC-adeno and CRPC-NE have dramatically distinct transcriptomes, including their splicing landscape, suggesting a critical role of splicing in PCa transdifferentiation.[14,15] Indeed, treatment-induced reshaping of alternative splicing (AS) pattern in PCa has been proposed to contribute to therapy resistance and cellular plasticity toward stemness.[15] Major spliceosome-mediated splice pattern shifts, including intron retention and/or isoform switching of PCa-relevant genes, have been extensively studied in prostate tumorigenesis and progression.[9,15–17] However, little is known about the pathways controlling non-canonical or minor intron splicing.

Here, we show that MIG expression and MiS activity are tightly linked to (prostate) cancer onset/progression and that AR signaling regulates MiS function as PCa progresses to CRPC-adeno. Minor intron splicing might regulate CRPC-adeno to CRPC-NE transition via the AS of RE1-silencing fact (REST). A crucial MiS component, U6atac snRNA, was elevated across PCa metastasis, thereby resulting in more efficient minor intron splicing with PCa progression. Thus, we downregulated U6atac, which decreased the growth of common cell line and patient-derived organoid models of advanced PCa. In fact, U6atac knockdown (KD) outperforms current therapeutics, such as the combination of enzalutamide with EZH2 inhibition. Our work elucidates a so far unexplored pathway, the MiS, as a point of entry for therapeutics against lethal PCa, a strategy that could extend to other cancer types.

RESULTS

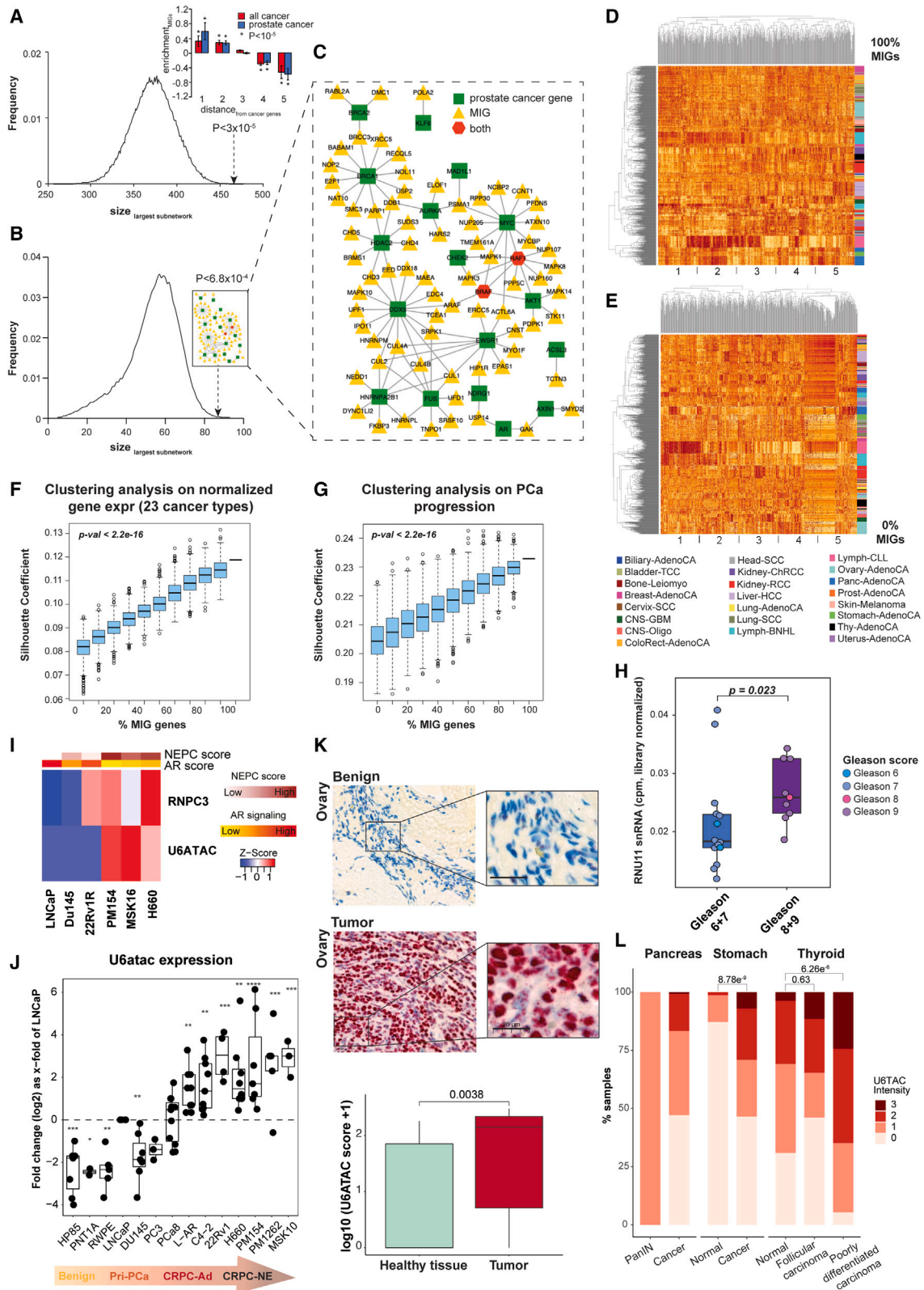### MIGs play a crucial role in cancer onset and progression
Because MIGs are enriched in cell cycle and survival,[18,19] we investigated the abundance of MIGs in molecular programs downstream of oncogenes. Analysis of the protein-protein interactome (PPI) of oncogenes showed that MIG-encoded proteins

were enriched among proteins that interact with proteins encoded by oncogenes. We computed the shortest distance d of all proteins encoded by oncogenes and found significant enrichment of MIGs in the immediate network neighborhood of oncogenes (permutation test: d = 1,2; p < 10E-5) and significant dilution at higher distances (d > 3, p < 10e−5) (inset, Figure 1A; STAR Methods). We observed significant enrichment of MIG-encoded proteins as immediate interactors of 26 verified PCa oncogenes (inset Figure 1A; Table S1).[20] In fact, MIGs were significantly (p < 3 × 10^{-5}) enriched in some of the largest PPI networks formed by oncogenes involved in all cancers (Figure 1A, right of the bell curve), and for 26 PCa oncogenes (permutation test, p < 6.8 × 10^{-4}). For example, 17 PCa oncogenes (hubs) were connected through 72 MIGs (spokes) (Figure 1B [inset] and 1C), suggesting that MIG-encoded proteins are tied to multiple oncogenic pathways.

MIGs expression alone could distinguish 23 different cancers (Figures 1D, 1E, and S1; Table S2). The silhouette score, which measures relative structure in heatmaps of hierarchical clustering, was highest when 100% MIGs were used to sort 23 cancer types (Figure 1F). However, a progressive reduction in the silhouette score was observed when non-MIGs were progressively added (Figures 1F and S1; Table S2; STAR Methods).[22–25] Similar enhanced performance in clustering of MIG-expression over non-MIGs was observed for differentiating PCa progression (Figures 1G, S2, and S3). These findings show that the transcription gene space of MIGs is closely associated with cancer onset and progression.

U6atac snRNA expression correlates with PCa progression
The upregulation of MIGs with PCa progression and transdifferentiation would necessitate increased MiS activity, which relies on U11, U12, U4atac, and U6atac snRNA expression. When we queried the expression kinetics of U11 snRNA (*RNU11*), a crucial MiS component, in PCa TCGA (prostate adenocarcinoma [PRAD]) data, we found a significant association between high-grade (Gleason score 8 and 9) PCa patients and high *RNU11* expression (Figure 1H; Table S3). Analysis of RNA sequencing (RNA-seq) data from 18 PCa organoids (CRPC-adeno and CRPC-NE) showed that expression RNPC3, a MiS component linked to cancer-associated scleroderma,[26,27] had a tendency toward positive correlation with *SOX2*, the pluripotent stem cell and early differentiation marker, and the epigenetic regulator *EZH2* (Figure S4A). *RNPC3* expression paralleled PCa disease progression with the lowest expression in benign and hormone-sensitive PCa cells (LNCaP), intermediate expression in aggressive CRPC-adeno cells (22Rv1), and the highest expression in CRPC-NE cells (H660) and patient-derived organoid lines (PM154 and MSK16) (Figure 1I). RNPC3 bridges U11/U12 di-snRNP, which tracked with upregulation of U11 and U12 snRNA in CRPC-adeno and CRPC-NE, and the same was observed for U4atac snRNA (Figure S4B). U6atac snRNA, which is normally maintained at lower levels through rapid degradation, is a bottleneck that throttles MiS activity. Thus, we hypothesized elevated U6atac expression across PCa progression. In agreement, U6atac levels tracked with high NEPC and low AR scores, which indicate poor PCa prognosis due to CRPC-NE lineage

plasticity (Figure 1I). Quantitative real-time PCR (qPCR) on cDNA from various cell lines and organoids confirmed that U6atac expression correlates with PCa progression: U6atac expression was higher in CRPC-adeno and CRPC-NE cell lines relative to normal HP85 prostate organoids, hormone-responsive LNCaP and RWPE cells, or to AR-low Du145 and PC3 cells (Figure 1J).

U6atac expression by *in situ* analysis showed a progressive increase across PCa progression (Figure S5). Similarly, U6atac *in situ* signals were higher in cancer tissue compared with matching benign samples of ovary (Figure 1K), small-cell lung cancer (SCLC), lung adenocarcinoma (ADC), skin, and uterus (Figure S5A). This finding demonstrates a strong specificity of U6atac expression for cancerous tissues (pan-cancer tissue microarray [TMA]) (Figures 1L and S5A). Consistently, thyroid and stomach cancer showed a higher U6atac intensity score in cancer compared with the corresponding benign tissues (Figures 1L and S5B). Thus, the expression of MiS components, especially U6atac, and MIGs shows a reliable association with cancer types and PCa progression.

MiS activity correlates with PCa progression

The high levels of U6atac in more aggressive PCa cell lines (Figures 1J–1L) were reflected in primary and metastatic PCa pa-tient samples compared with benign prostate (Figures 2A and S6A). We tested the hypothesis that MiS activity increases with PCa progression by employing minor and major splicing reporter plasmids[28] in cell lines reflecting PCa progression and transdif-ferentiation. Minor intron-splicing efficiency was higher in ther-apy-resistant C4-2 and 22Rv1 cells compared with benign HP85 prostate organoids, PNT1A cells, and hormone-sensitive LNCaP cells (Figure 2B, upper panel). Similarly, MiS efficiency was higher in advanced CRPC-adeno and CRPC-NE models compared with AR-low DU145 and PC3 cells (Figure 2B, upper panel). We also observed a 6-fold increase in minor intron splicing in the CRPC-NE cells and patient-derived organoid lines (PM154,[29] PM1262,[29] MSK16,[30] and NCI-H660[31]) (Figure 2B, upper panel). By contrast, major intron splicing was unaffected in all cell and organoid lines tested (Figure 2B, lower panel), sup-porting the selective increase in MiS activity with PCa progres-sion. Many MIGs were expressed in CRPC-NE PM154 compared with therapy-sensitive LNCaP cells (Figure S6B; Table S4). Together, these data suggest an increase in MiS requirement with PCa progression (Figure 2B).

Whereas the upregulated MIGs in therapy-responsive LNCaPs enriched for anti-viral response, those in C4-2 cells en-compassed DNA damage and autophagy (Figure S6B). MIGs of

Figure 1. MIG expression patterns and MiS component expression correlate with PCa progression

(A) Enrichment of MIGs (542) among proteins interacting at a distance (d) = 1, d = 2, d = 3, d = 4, and d = 5, with proteins encoded by 403 cancer and 26 PCa genes in a human protein-protein interaction network. Enrichment values show the $\log_2$ fold change of the presence of MIGs in distance bins, compared with randomly sampled gene sets of equal size. Error bars indicate standard deviations. Asterisk indicates p < 10e−5.

(B) Largest connected subnetwork of all interactions between PCa genes and MIGs captured 87 genes (inset). Curve indicates the distribution of network sizes when we randomly sampled sets of non-MIGs of equal size (542).

(C) Subnetwork of interactions between 74 MIGs and 19 PCa genes (amplification of inset in B).

(D) Heatmap gene expression data from 23 cancer types totaling n = 1,224 samples. Gene expression data were taken from PCAWG. Clustering was performed using an n_cluster value of K = 4, wherein 100% of the genes are MIGs.

(E) Heatmap gene expression data from 23 cancer types totaling n=1,224 samples. Gene expression data were taken from PCAWG. Clustering was performed using an n_cluster value of K = 4, wherein 100% of the genes are non-MIGs.

(F) Relative performance of MIGs and non-MIGs with respect to clustering gene expression data from 23 distinct cancer types from PCAWG. Each box cor-responds to a 1,000-length simulation in which non-MIGs are randomly sampled from among all non-MIGs in the genome. This plot exhibits a negative rela-tionship between the relative abundance of non-MIGs and the quality of clustering, as quantified using the silhouette coefficient. The p value is based on a two-sided t test in which each sample value is taken to be the difference between the silhouette coefficient of one of the 1,000 100% non-MIG samples and the corresponding silhouette coefficient when only MIGs are used for clustering (i.e., the silhouette coefficient corresponding to 0% non-MIGs).

(G) Relative performance of MIGs and non-MIGs with respect to clustering gene expression data from different stages of PCa progression. Gene expression data were taken from GTEx, TCGA, and SU2C, respectively, and these datasets represent increasing stages of tumor progression (from healthy tissue in GTEx to advanced prostate tumors in SU2C). Clustering was performed using an n_cluster value of K = 3. Each box in this plot corresponds to a 1,000-length simulation in which non-MIGs are randomly sampled from among all non-MIGs in the genome. This plot exhibits a negative relationship between the relative abundance of non-MIGs and the quality of clustering, as quantified using the silhouette coefficient. The p value is based on a two-sided t test in which each sample value is taken to be the difference between the silhouette coefficient of one of the 1,000 100% non-MIG samples and the corresponding silhouette coefficient when only MIGs are used for clustering (i.e., the silhouette coefficient corresponding to 0% non-MIGs).

(H) Boxplots representing the distribution of *RNU11* expression values (counts per million [CPM]) across primary tumor samples where *RNU11* was detected. Samples (n = 23) were grouped according to low (6 + 7) or high (8 + 9) Gleason scores (n = 23, Wilcoxon test, p = 0.023).

(I) Heatmap showing RNA-seq expression (FPKM) of PCa cell lines, ordered by increasing NEPC and decreasing AR score. The NEPC and AR scores were calculated based on FPKM values of a set of 70 and 27 genes to estimate the likelihood of a test sample to be CRPC-NE or CRPC-adeno, respectively.[21]

(J) U6atac snRNA expression as x fold of siScrambled normalized to the mean of GAPDH and ACTB gene transcription in different PCa cell lines (HP85 n = 5, PNT1A n = 2, RWPE n = 5, LNCaP n = 7, DU145 n = 7, PC3 n = 3, PCc8 n = 9, L-AR n = 10, C4-2 n = 7, 22Rv1 n = 5, PM154 n = 5, H660 n = 5). Data are represented using box and whisker plots. Boxplots display values of minimum, first quartile, median, third quartile, and maximum (two-sided Mann-Whitney test, ns; not significant, p > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001). Each data point represents the x fold of LNCaP cells from a single experiment. Experiments were performed in triplicates.

(K) Representative images of the validation of RNU6atac BaseScope probes in human ovary benign and tumor tissue. Scale bars represent 50 μm. Lower panel shows the U6atac score of the TMA analysis (Wilcoxon test, p = 0.0038).

(L) U6atac score analysis of human PanIN (pancreatic intraepithelial neoplasia) benign (n = 1) and tumor (n = 192) tissue, stomach benign (n = 82) and tumor (n = 102) tissue, and from thyroid benign tissue (n = 142), thyroid follicular carcinoma (n = 26), and poorly differentiated thyroid carcinoma (n = 39) (Wilcoxon test, stomach: p = 8.78e−9, thyroid follicular carcinoma: p = 0.63, thyroid poorly differentiated carcinoma: p = 6.26e−6).

See also Figures S1–S5 and Tables S1, S2, and S3.
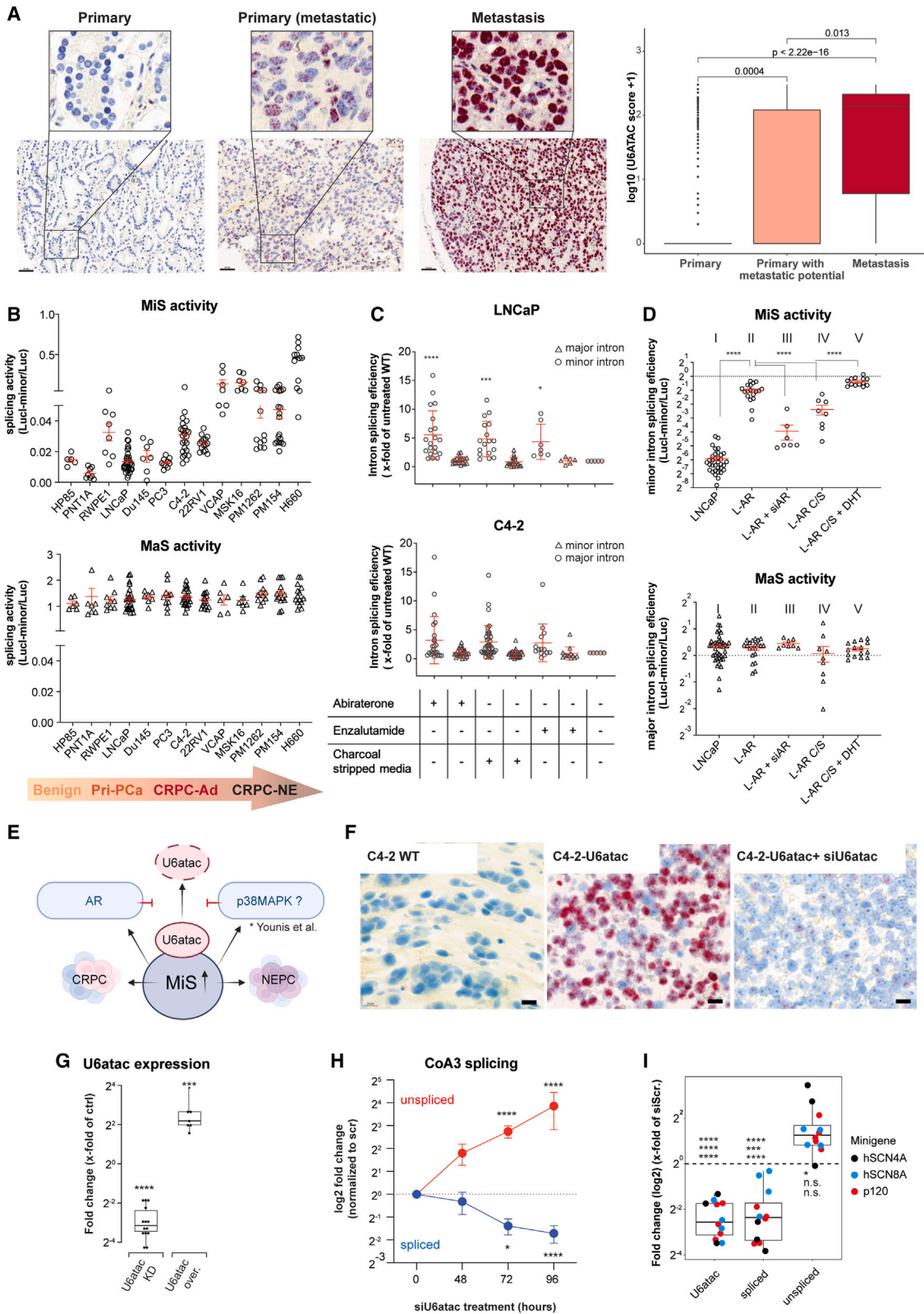
(legend on next page)

22Rv1 included MAPK11 and tumor antigens (CTAG2) or splice factors (SRPK3), but those upregulated in PM154 organoids included MAPK10, chromatin remodeling, cytoskeletal signaling (CEP170, EML4, and PDE6D), and DNA damage repair (MSH3) (Figure S6B). We also observed unique AS events across minor introns of MIGs, such as *TSPYL2*, which contribute to arbiraterone, an ARSi, resistance in CRPC-adeno,[32] in all four cell lines, with higher levels often in PM154 (Figure S6B). Thus, various sets of MIGs are leveraged for PCa progression.

Because AR signaling plays a critical role in PCa progression to CRPC-adeno, we explored whether AR signaling regulates MiS activity in PCa. To simulate stress response and re-activation of AR signaling in PCa, we subjected PCa cells to long-term ADT and ARSi using charcoal-stripped (c/s) media, abiraterone, and enzalutamide. We observed a significant increase in MiS activity in cells exposed to ADT/ARSi, whereas the treatment had only limited effects on major splicing (Figure 2C). A similar association was observed in PC3 and PC3Pro4, as well as in hormone-sensitive LNCaP cells and an LNCaP (L) subline, L-AR and/or is enzalutamide-resistant (L-rENZ) (Figures S6C and S6D). This increase in MiS activity in those lines was further confirmed by the minor intron-containing p120 and hSCN4A minigene constructs (Figure S6E). Lastly, L-rENZ and PC3Pro4 cells expressed higher MiS snRNA levels (Figure S6F) than their respective wild-type (WT) lines.

Next, we used the luciferase minor intron-splicing reporter to study the relationship of AR signaling to minor intron-splicing therapy-sensitive LNCaP cells. The overexpression of AR led to a significant increase in MiS activity (Figure 2D II), which was not observed with the major spliceosome reporter (Figure 2DII, lower panel). MiS activity decreased upon siRNA-mediated downregulation of AR (Figure 2DIII) confirmed by qRT-PCR (Figure S7A). We next treated L-AR cells with hormone-depleted media to block AR activation, which was confirmed by a reduction in *KLK3* expression (Figure S7B), resulting in decreased MiS activity (Figure 2DIV). Finally, adding dihydrotestosterone (DHT) to the hormone-depleted media rescued MiS activity to levels observed in L-AR alone (Figure 2DV). Similar findings were observed with the p120 minor intron-splicing minigene reporter (Figure S7C). AR modulation did not affect major spliceosome activity (Figure 2D, lower panel). When we tested whether AR modulation regulates MiS activity through U6atac snRNA levels,[33] we found that modulation of AR levels and activity does result in changes in U6atac snRNA levels (Figures S6F and S7D).

Unlike CRPC-adeno, MiS activity in CRPC-NE, which lacks AR expression, must be through an AR-independent pathway, such as the p38MAPKs (MAPK11,12,13, and 14), which are themselves MIGs (Figure 2E).[33] Analysis of our U6atac KD RNA-seq and proteomics datasets (Figures 3 and 4; Tables S5 and S6) revealed that p38MAPK family members display intron retention or AS events in LNCaP, C4-2, 22Rv1 cells, and the PM154 organoid. We found that MAPK14 protein expression is only decreased in 22Rv1 cells, representing an intermediate between adeno and NE states. MAPK13, which promotes neurotoxicity and is required for prostate epithelial differentiation,[34,35] was only decreased in PM154 organoids (Table S5C).

Figure 2. MiS activity is regulated by AR and is elevated in PCa

(A) Representative images of the validation of RNU6atac BaseScope probes in human primary (325 cases), primary with metastatic potential (25 cases), and metastatic (32 cases) PCa tissue. Scale bars represent 50 μm. Quantification of the U6atac score of the whole TMA analysis shown on the right-hand side. Comparisons between groups were performed with Wilcoxon test (primary vs. primary (metastatic: p = 0.004, primary vs. metastasis: p < 2.22e−−16, primary (metastatic) vs. metastasis: p = 0.013).

(B) Normalized luminescence values of minor/major spliceosome luc-reporter plasmids in different normal prostate and PCa cell lines of different PCa subtypes (HP85 n = 5, PNt1A n = 7, RWPE n = 9, LNCaP n = 30, DU145 n = 7, PC3 n = 10, C4-2 n = 22, 22Rv1 n = 13, VCAP n = 7, PM1262 n = 13, PM154 n = 20, MSK16 n = 7, H660 = 12, mean ± SEM, one-way ANOVA; ns p > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001). Each data point represents a single experiment; experiments were performed in triplicates. Abbreviations: Pri-PCa, primary PCa; CRPC-Ad, castration-resistant adeno PCa, CRPC-NE, castration-resistant neuroendocrine PCa.

(C) Normalized luminescence values of minor/major spliceosome luc-reporter plasmids in LNCaP and C4-2 PCa cells subjected to long-term treatment with abiraterone, enzalutamide, and charcoal stripped (C/S) media (LNCaP: abiraterone n = 19, c/s n = 17, enzalutamide n = 7, C4-2: abiraterone n = 24, c/s n = 29, enzalutamide n = 12, mean ± SEM, one-way ANOVA; ns p > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001). Each data point represents a single experiment; experiments were performed in triplicates.

(D) Normalized luminescence values of minor/major spliceosome luc-reporter plasmids in LNCaP and L-AR cells treated with siAR, c/s media, and c/s media + DHT (10 μM) (LNCaP n = 34, L-AR n = 19, L-AR + siAR n = 7, L-AR c/s n = 8, L-AR c/s + DHT n = 13, mean ± SEM, ordinary one-way ANOVA; ns p > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001). Each data point represents a single experiment, each experiment was performed in triplicate. Abbreviations: c/s, charcoal-stripped media; DHT, dihydrotestosterone.

(E) Schematic overview of U6atc turnover regulation. AR blocks U6atac turnover in CRPC, which leads to increased MiS activity, whereas p38 does so in NEPC.
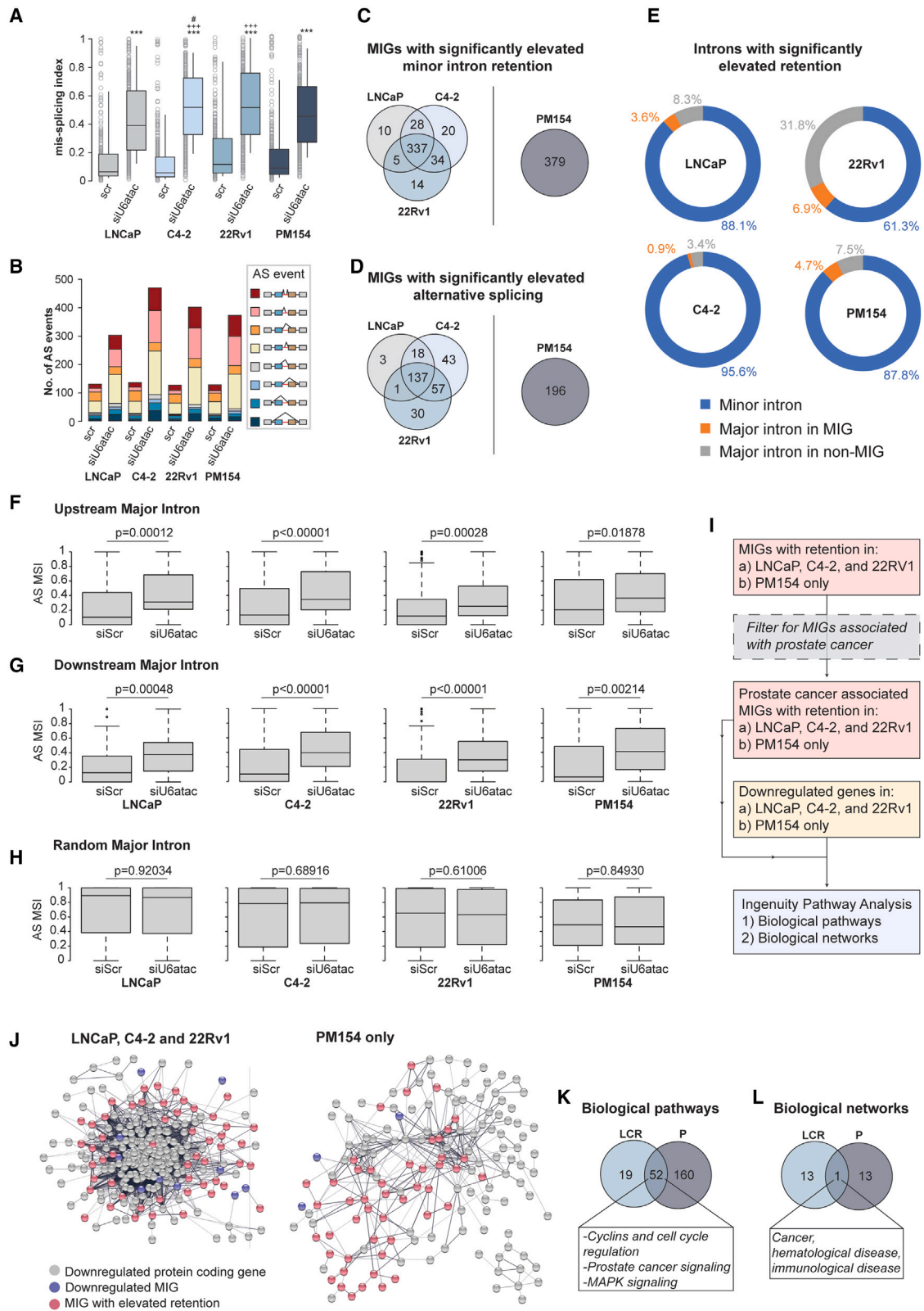
(F) Representative images of the validation of RNU6atac BaseScope probes in C4-2 wild-type (WT) cells, in C4-2 cells that overexpress U6atac (C4-2-U6atac) and in C4-2 cells overexpressing U6atac that were treated with siU6atac RNA for 96 h. Scale bars represent 20 μm.

(G) U6atac snRNA expression as x fold of siScrambled (control) normalized to the mean of GAPDH and ACTB gene transcription in C4-2 cells (U6atac knockdown n = 14, U6atac over. n = 7) Data are represented using box- and whisker plots that display values of minimum, first quartile, median, third quartile, and maximum. Statistical analysis was evaluated using a two-sided unpaired t test, ns p > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001. Each data point represents a single experiment; experiments were performed in triplicates.

(H) Expression of *CoA3* spliced and unspliced transcripts as x-fold of siScrambled normalized to the mean of GAPDH and ACTB gene transcription in C4-2 cells treated with siU6atac RNA for 48, 72, and 96 h (n = 3, two-way ANOVA; ns p > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001). Experiments were performed in triplicates.

(I) U6atac and hSCN4A, hSCN8A, p120 spliced and unspliced expression as x fold of siScrambled (siScr.) normalized to the mean of GAPDH and ACTB gene transcription in C4-2 cells transfected with the hSCN4A, hSCN8a, and p129 minigene reporter plasmids and treated with siU6atac RNA for 96 h (hSCN4A n = 3, hSCN8A n = 4, p120 n = 5). Data are represented using box and whisker plots. Boxplots display values of minimum, first quartile, median, third quartile, and maximum (two-sided unpaired t test, ns p > 0.05, ***p < 0.001, ****p < 0.0001). Each data point represents a single experiment, experiments were performed in triplicates.

See also Figures S6 and S7.

A

B

C MIGs with significantly elevated minor intron retention

D MIGs with significantly elevated alternative splicing

E Introns with significantly elevated retention

- Minor intron
- Major intron in MIG
- Major intron in non-MIG

F Upstream Major Intron

G Downstream Major Intron

H Random Major Intron

I
- MIGs with retention in:
  a) LNCaP, C4-2, and 22RV1
  b) PM154 only
- Filter for MIGs associated with prostate cancer
- Prostate cancer associated MIGs with retention in:
  a) LNCaP, C4-2, and 22Rv1
  b) PM154 only
- Downregulated genes in:
  a) LNCaP, C4-2, and 22Rv1
  b) PM154 only
- Ingenuity Pathway Analysis
  1) Biological pathways
  2) Biological networks

J LNCaP, C4-2 and 22Rv1 / PM154 only

- Downregulated protein coding gene
- Downregulated MIG
- MIG with elevated retention

K Biological pathways

- Cyclins and cell cycle regulation
- Prostate cancer signaling
- MAPK signaling

L Biological networks

- Cancer, hematological disease, immunological disease

(legend on next page)

MiS inhibition in PCa results in aberrant minor intron splicing

Because MiS plays a crucial role in PCa progression, we sought to inhibit the MiS in PCa cell lines (LNCaP, C4-2) and the patient-derived organoid (PM154) representing disease progression. We employed RNAi to KD U6atac and MiS activity.[36–47] The efficacy of the RNAi was tested by treating C4-2 cells overexpressing U6atac with siU6atac (Figures 2F and S5C) and the effects on minor splicing were examined by monitoring the splicing of the minor intron of COA3 and minigene reporters[33] (Figures 2G–2I and S7E). Similar MiS inhibition was observed in siU6atac-treated 22Rv1, PM154, L-AR, LNCaP, L-rENZ, and H660 cells and the organoid (Figure S7F).

After confirming equivalent levels of U6atac KD in siU6atac versus siScrambled-treated cell lines (LNCaP, C4-2, and 22Rv1) and organoid (PM154) for 96 h (Figure S8A), MiS inhibition was captured by ribo-depleted total RNA-seq. Quantifying the mis-splicing index for each sample revealed significantly elevated minor intron retention in siU6atac-treated LNCaP, C4-2, 22Rv1 cells, and PM154 compared with their respective siScrambled control (Figure 3A; Tables S6A–S6E). AS events, especially the use of cryptic SSs across minor intron, were the most pronounced effect of siU6atac treatment in C4-2 cells, followed by 22Rv1 cells, then PM154, and lastly, LNCaP cells (Figure 3B; Tables S6F–S6J).

We found significant elevation of minor intron retention in 337 and 389 MIGs common to all cell lines, and PM154-organoid, respectively (Figure 3C; Table S6E). MIGs with elevated retention enriched for gene ontology (GO)-terms such as RNA and DNA processing, vesicle transport, and mRNA splicing (Table S7). Elevated AS events in 137 and 196 MIGs was common to all

cell lines (Figures 3C and 3D; Table S6J) and PM154-organoid, respectively (student's two-tailed t test, p < 0.05) (Table S6J). The number of MIGs with elevated minor intron retention is highest in LNCaP > C4-2 > 22Rv1 > PM154 (Figure 3E). Few major introns in non-MIGs showed elevated retention in LNCaP, C4-2, and PM154 models, except for 22Rv1, thereby confirming the specificity of siU6atac-mediated MiS inhibition, which confirmed the specificity of siU6atac (Figures 3H, S8B, and S8C; Table S6K). By contrast, major introns flanking minor introns in MIGs showed statistically elevated retention in siU6atac-treated cells (Figures 3F, 3G, S8B, and S8C).

Aberrant minor intron splicing in MIGs with disparate functions was anticipated to affect the overall transcriptome of PCa, which we captured by differential gene expression analysis (Figure S9). The shared downregulated genes enriched for many GO terms related to the cell cycle (Table S7). Alternatively, the 189 downregulated genes in 96 h siU6atac-treated PM154 cells enriched for a single GO term—cell differentiation—which was unique to this organoid (Tables S7 and S8). Gene set enrichment analysis (GSEA) based on hallmark gene sets on the up- and downregulated genes of all three cell lines and organoid confirmed that pro-proliferative (E2F targets, G2M checkpoint, and mitotic spindle) and DNA repair pathways were reduced upon siU6atac-mediated MiS inhibition and, additionally, revealed a reduction in prostate-specific pathways, such as androgen response or spermatogenesis (Figure S10).

Next, we integrated MIGs with minor intron retention common to LNCaP, C4-2, and 22Rv1 cell lines (Figure 3I) and identified direct interactors of PCa-causing genes (Figures 1B and 1C), which yielded 55 MIGs for the LNCaP, C4-2, and 22RV1 cell lines (Table S9). Most downregulated genes formed a tight-knit

Figure 3. Inhibition of the MiS through U6atac siRNA effectively alters the PCa transcriptome

(A) Boxplot showing 10th–90th percentile mis-splicing index for all minor introns that show retention (431) in LNCaP, C4-2, and 22Rv1 cell lines, as well as PM154 organoid after 96 h treatment with siScrambled (scr) or siU6atac. Significance determined by Kruskal-Wallis test followed by post hoc Dunn's test: asterisk (*) denotes significance compared with each siU6atac samples' appropriate siScrambled control; plus sign (+) denotes significance compared with LNCaP siU6atac; hashtag (#) denotes significance compared with PM154 siU6atac. *p < 0.05, **p < 0.01, ***p < 0.001.

(B) Bar chart showing the distribution of alternative splicing (AS) events around (from two exons upstream to two exons downstream) all minor introns in LNCaP, C4-2, and 22Rv1 cell lines and the PM154 organoid after 96 h treatment with siScrambled (scr) or siU6atac. Different colors depict different alternative splicing categories.

(C) Venn-diagram (not to scale) depicting the overlap of MIGs with significantly elevated minor intron retention in the LNCaP, C4-2, and 22RV1 cell lines (left) and the PM154 organoid (right) after 96 h siU6atac treatment compared with the appropriate 96 h siScrambled control.

(D) Venn-diagram (not to scale) depicting the overlap of MIGs with significantly elevated alternative splicing in the LNCaP, C4-2, and 22RV1 cell lines (left) and the PM154 organoid (right) after 96 h siU6atac treatment compared with the appropriate 96 h siScrambled control.

(E) Cake- diagram displaying results from an IRFinder intron retention analysis for each cell line. Intron retention was considered significantly elevated if $\log_2 FC > 1$ and BH-adjusted p < 0.05.

(F) Analysis of AS mis-splicing index (MSI) for all AS events occurring in major introns upstream of a minor intron in either the siScrambled (siScr) or siU6atac condition. Data projected for each cell line as boxplots (Mann-Whitney U test with p < 0.05).

(G) Analysis of AS mis-splicing index (MSI) for all AS events occurring in major introns downstream of a minor intron in either the siScrambled (siScr) or siU6atac condition. Data projected for each cell line as boxplots (Mann-Whitney U test with p < 0.05).

(H) Analysis of AS mis-splicing index (MSI) for all AS events occurring in random major introns in either the siScrambled (siScr) or siU6atac condition. Data projected for each cell line as boxplots (Mann-Whitney U test with p < 0.05).

(I) Pipeline for generating and analyzing interaction networks between shared (i) MIGs with significantly elevated minor intron retention and (ii) significantly downregulated protein-coding genes in all three cell lines (a) or in the PM154 organoid (b). MIGs associated with PCa are shown in Figure 1C.

(J) STRING network showing association of shared (i) PCa-associated MIGs with elevated minor intron retention (red) and (ii) downregulated protein-coding genes (gray; downregulated MIGs in blue) in all three cell lines LNCaP, C4-2, and 22RV1 (left) and PM154 organoid (right).

(K) Venn-diagram (not to scale) showing the overlap of ingenuity pathway analysis (IPA)-generated biological pathways for the LNCaP, C4-2, and 22RV1 (LCR) gene list (H; left) and PM154 (P) gene list (H; right).

(L) Venn-diagram (not to scale) showing the overlap of ingenuity pathway analysis (IPA)-generated biological networks for the LNCaP, C4-2, and 22RV1 (LCR) gene list (H; left) and PM154 (P) gene list (H; right).

See also Figures S8–S10 and Tables S4, S5, S6, S7, S8, S9, and S10.
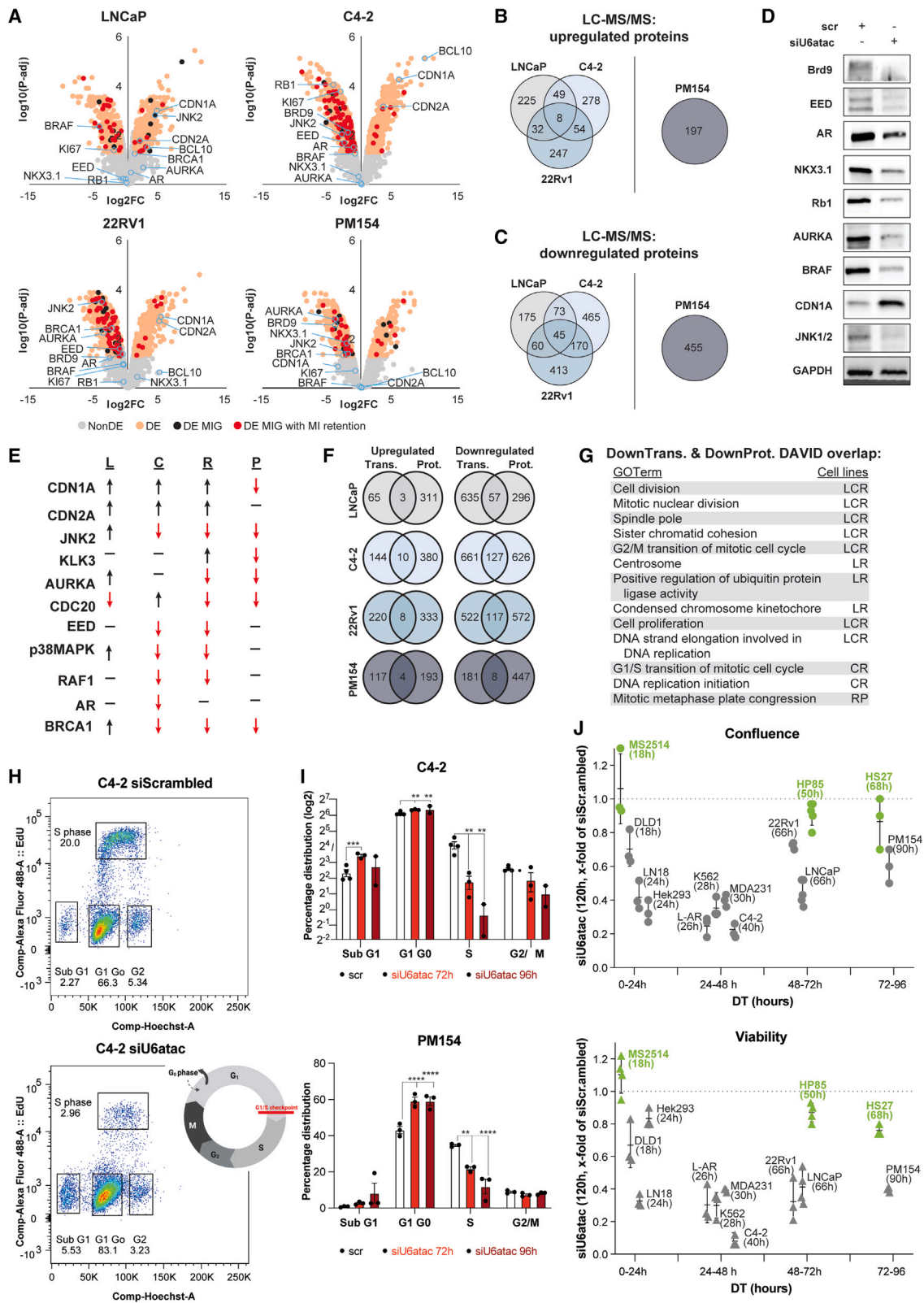
(legend on next page)

central network surrounded by MIGs with high minor intron retention (Figure 3J). Ingenuity pathway analysis (IPA) revealed 52 common biological pathways underlying these networks, such as cell-cycle regulation, PCa signaling, apoptosis, and MAPK signaling (Figures 3K and 3L; Table S10). Common subnetworks enriched in the three cell lines were cell cycle and DNA damage response (DDR). The organoid sub-networks associated with oncogenic signaling pathways and with actions executed by the cell cytoskeleton. Overall, siRNA-mediated downregulation of U6atac resulted in a robust minor intron-splicing defect in a large subset of MIGs involved in cancer-relevant pathways.

Impact of MiS inhibition on the PCa proteome

The effect of minor intron-splicing defect on the up/downregulation of proteins was captured by LC-MS/MS on the same cell lines and organoids treated with siRNA (Figures 4A–4C and S11A). We confirmed changes in levels observed by LC-MS/MS for several PCa-relevant proteins by immunoblot (Figure 4D). Among others, the AR (non-MIG), Rb1 (non-MIG), the epigenetic regulator EED (MIG) and common target during PCa therapy, JNK1 (MIG), which is another potential PCa target, and the proliferation marker Ki67 (non-MIG) were strongly decreased upon siU6atac treatment in the CRPC-adeno line C4-2 (Figures 4D, 4E, S11B, S11C; Table S11). Similarly, we observed downregulation of both PARP (MIG) and EZH2 (non-MIG) proteins with increasing siU6atac treatment time (Figure S12A).

Comparing the proteomic data of the four analyzed cell lines representing PCa disease progression revealed a cell type and, thus, probably a PCa subtype and context-specific MiS-dependent proteome very similar to the RNA-seq analysis. Each cell line expressed a unique set of up and downregulated proteins, including MIGs and non-MIGs, (Figures 4B, 4E, S11B, and S11C; Tables S5 and S11), each critical for cancer biology. There was discordance between the transcriptome and proteome data

in each cell line (Figure 4F), which is consistent with reports that some aberrantly spliced MIGs can escape non-sense-mediated decay (NMD).[48] For example, we found that MIG-encoded proteins, such as EED, JNK1/2, BRAF, and RAF1, decreased, although their expression by RNA-seq remained unchanged, albeit with high minor intron retention and AS upon siU6atac treatment (Figures 4A–4E, S11B, and S11C; Table S11). However, overall analysis of the MIG transcripts showed that the most common pattern was elevated retention, predicted NMD, and downregulated protein (Figure S12B; Table S12F).

We also found non-MIG-encoded proteins, such as the AR and Rb1, to be decreased only in the proteome, implying indirect regulation through the MiS (Figures S11B, S11C, and S12A). Despite this, we identified genes whose mRNA transcripts and encoded proteins were significantly downregulated after 96 h of siU6atac treatement (Figure 4F; Table S12). These four gene sets showed a high degree of GO Term overlap that centered on cell cycle (Figure 4G). Taken together, siU6atac-mediated MiS inhibition altered the transcriptome and the proteome, which block cell cycle and survival of PCa.

This implication of MiS inhibition on cell-cycle progression led us to employ FACS analysis at 72- and 96-h post-transfection (Figures S12C and S12D). There was a significant increase in G1/G0 phase cells and a significant decrease in S-phase cells in therapy-resistant C4-2 and PM154 organoids cells treated with siU6atac (Figures 4H, 4I, and S12C). The cell-cycle defects that were observed revealed that MiS inhibition provokes a G1/S cell-cycle arrest in PCa. However, when we investigated the relationship between proliferation rate and MiS inhibition in normal and tumor cells (Figures 4J and S12E), we did not observe a correlation between cell growth (doubling time) and siU6atac response. Therefore, MiS inhibition, although affecting cell cycle and proliferation, is not correlated with the proliferation rates of the cells. The extent of MiS dependency is cancer type and stage specific. Indeed, we also show that PCa-specific oncogenic

Figure 4. U6atac-mediated MiS inhibition effectively alters the PCa proteome in a cell type-specific manner

(A) Volcano plot showing proteins most significantly increased (upper right) and decreased (upper left) in C4-2 cells treated with siU6atac (96 h), compared with the siScrambled control (pooled data from 3 coIP replicates). The x axis represents $\log_2$ fold change (FC) values, the y axis represents $-\log_{10}$ of adjusted p values. Gray dots represent non-differentially expressed (non-DE) proteins; orange dots represent differentially expressed proteins (DE), black dots represent differentially expressed MIG-encoded proteins (DE MIG) and red dots represent differentially expressed MIG-encoded proteins that transcriptomically have significantly elevated minor intron retention.

(B) Venn diagrams (not to scale) illustrating the overlap in proteins that are upregulated after U6atac knockdown in LNCaP, C4-2, 22Rv1, and PM154 cells, assessed by mass spectrometry analysis.

(C) Venn diagrams (not to scale) illustrating the overlap in proteins that are downregulated after U6atac knockdown in LNCaP, C4-2, 22Rv1, and Pm154 cells, assessed by mass spectrometry analysis.

(D) Immunoblot showing expression levels of selected proteins, which are down- or upregulated in mass spectrometry analysis in C4-2 cell lines treated for 96 h with siScrambled or siU6atac RNA. GAPDH was used as loading control. Image shows a composite of a membrane that was cropped to only show C4-2 cells as an exemplar. Original gel image containing LNCaP, C42, 22Rv1, and Pm154 samples is published through Mendeley data.

(E) Table summarizing differential expression of selected proteins after siU6atac treatment in LNCaP (L), C4-2 (C), 22Rv1 (R), and PM154 (P) cells.

(F) Venn-diagrams (not to scale) illustrating the overlap of genes that show upregulation (left) or downregulation (right) transcriptionally (trans.) and by mass spec (prot.) in LNCaP, C4-2, 22Rv1, and PM154 cells.

(G) Table summarizing the shared DAVID GO Terms for the genes that are downregulated both transcriptionally (DownTrans.) and by mass spec (DownProt.) for each cell line. L = LnCAP, C = C4-2, R = 22Rv1, P = PM154.

(H) Psuedocolor plot showing the gates for G1/0, S, and G2 selection in C4-2 cells treated for 72 h with siU6atac (bottom) or siScrambled (top).

(I) Quantification of flow cytometry analysis of C4-2 and PM154 organoids treated with siU6atac for 72 h and 96 h, followed by staining with EdU and Hoechst (n = 3, mean ± SEM, ordinary two-way ANOVA; **p = 0.001, ****p < 0.001).

(J) Top: confluence (y axis) and bottom: viability (y axis) of different cell lines treated with siU6atac/siScrambled for 120 h (n = 3, mean + SD). Cells are ranked according to their doubling time (DT). Normal cells are depicted in green.

See also Figures S11 and S12 and Tables S5, S11, and S12.
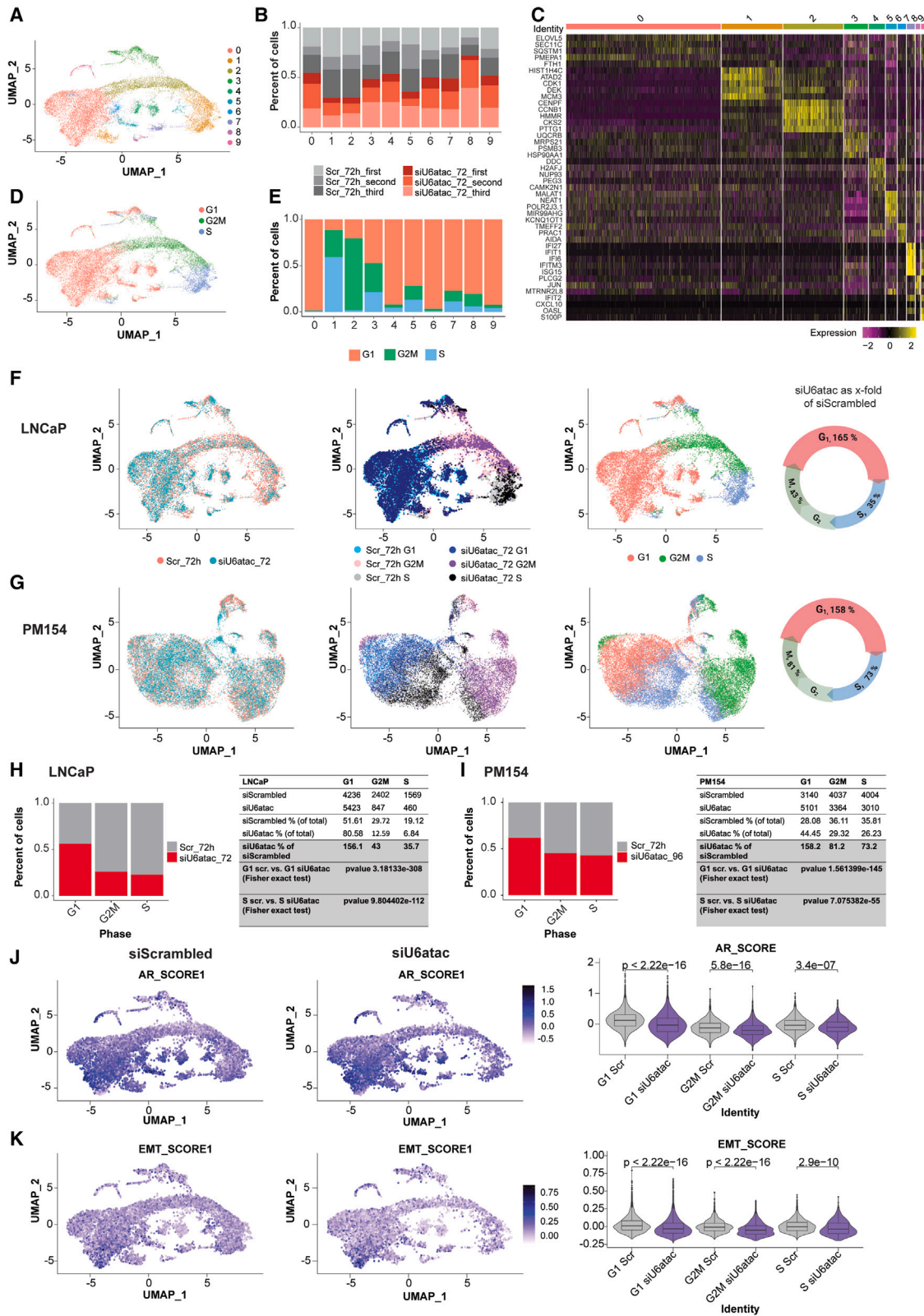
(legend on next page)

drivers, such as AR, are intertwined with MiS activity and might play a role in defining MiS dependency (Figure 2). We conclude that MiS inhibition disproportionally affects PCa cells over normal prostate cells.

Single-cell RNA-seq reveals cell-cycle defects triggered by MiS inhibition

To identify cellular heterogeneity in response to siU6atac-mediated MiS inhibition, we performed single-cell RNA sequencing (scRNA-seq) on LNCaP cells and PM154 organoids post siU6atac. After standard data processing and quality control procedures (STAR Methods), we obtained transcriptomic profiles for 8,206 siScrambled and 6,730 LNCaP cells and 11,181 siScrambled and 11,475 PM154 organoids (Figures S13A and S13B; Table S13). We employed unsupervised clustering to identify heterogeneity in response to siU6atac. K-means clustering and uniform manifold approximation and projection (UMAP) of the combined data across both genotypes revealed nine major cell clusters (Figure 5A). We found that clusters 0, 3, 4, 6, 7, and 9 were populated by an equal number of cells treated with siScrambled and siU6atac (Figure 5B). Clusters 1, (Fisher's exact test, $p = 3.17e-69$), 2 (Fisher's exact test, $p = 1.70e-63$), and 5 (Fisher's exact test, $p = 3.98e-05$) showed higher percentages of siScrambled cells, relative to siU6atac cells. Finally, cluster 8 (Fisher's exact test, $p = 1.53e-14$) was predominantly populated by cells treated with siU6atac compared with siScrambled (Figure 5B). GO term enrichment analysis of the top 5 upregulated genes in the cells belonging to clusters 1, 2, and 5 revealed cell division, G1/S transition, and mitotic nuclear division (Figure 5C). By contrast, there was no GO Term enrichment for the top 5 genes in cluster 8 (Figure 5C). The enrichment of cell-cycle terms combined with previous findings that MiS inhibition affects cell-cycle regulation in cancer (Figures 3 and 4) led us to employ a list of cell-cycle regulators as a means of sorting single cells based on cell-cycle stage.[49] We superimposed cell-cycle stage on the UMAP representation of the unsupervised data clustering (Figure 5D). We found that clusters 0, 3, 4, 6, 7, and 9 showed the highest percentage of G1 cells (Figure 5E). By contrast, clusters 1 and 2 showed the highest percentage of cells in S and G2/M phase, respectively (Figure 5E). Similar analysis

for PM154 organoids yielded 7 clusters (Figure S13C). A majority of the cells populating cluster 5 (Fisher's exact test, $p = 3.29e-40$) were from siScrambled, whereas cluster 6 (Fisher's exact test, $2.74e-57$) was predominantly populated by siU6atac-treated cells (Figure S13D). There was no Go Term enrichment for genes upregulated in those clusters. Yet, cluster 5 (Figure S13E), as with clusters 1, 2, and 5 of LNCaP cells, showed a high percentage of cells in S and G2/M phase, respectively (Figures S13F and S13G).

The comparison of the distribution of cells from each condition in the 9 clusters and the distribution of cells in the different phases of cell cycle led us to investigate cell-cycle defects through scRNA-seq analysis (Figures 5F–5I). Therefore, we superimposed the cell identities (i.e., siScrambled or siU6atac) onto the UMAP of the clusters along with the cell-cycle phases (Figures 5F–5I). For LNCaP cells, we found a significant enrichment (Fisher's exact test, $p = 3.18e-308$, OR = 3.89, 95% CI 3.60–4.19) of siU6atac-treated cells in G1 phase and a reduction of these cells in S phase (Fisher's exact test, $p = 9.80e-112$, OR = 3.22, 95% CI 2.88–3.60) (Figures 5F and 5H). Similarly, for PM154 organoids, we observed a significant enrichment of siU6atac cells in G1 (Fisher's exact test, $p = 1.56e-145$, OR = 2.05, 95% CI 1.93–2.16) and a decrease in S phase (Fisher's exact test, $p = 7.07e-55$, OR = 1.57, 95% CI 1.48–1.66) (Figures 5G and 5I). These findings are consistent with FACS analysis (Figures 4H, 4I, and S12C).

Given that siU6atac is successful at blocking PCa progression and that AR signaling is a crucial driver of PCa, we next explored the AR score (a metric of CRPC-adeno progression) in siU6atac-treated LNCaP cells and PM154 organoids (Figures 5J and S13H). We discovered that siU6atac was, indeed, able to reduce the AR scores of both models significantly in G1 (Wilcoxon test, $p < 2.22e-16$), G2M (Wilcoxon test, LNCaP: $p = 5.8e-16$, PM154: $p < 2.22e-16$), and S phase (Wilcoxon test, LNCaP: $p = 3.4e-07$, PM154: $p < 2.22e-16$). We did not observe any shift in the AR score post siU6atac treatment in bulk RNA-seq (Figure S13J), which is not surprising, because we lose single-cell resolution and dilution of expression of some of the critical AR-responsive genes. Metastasis is often measured by endothelial to mesenchymal transition (EMT) score.[50] We found that

Figure 5. Single-cell RNA-seq corroborates siU6atac-mediated cell-cycle defects and reveals PCa lineage dependency on MiS function
(A) UMAP representation of LNCaP cell line showing the clusters at the optimal resolution (0.2).
(B) Histogram of LNCaP cell line showing the contribution of each sample in each cluster.
(C) Heatmap of LNCaP cell line showing the expression of top 5 marker genes in each cluster.
(D) UMAP representation of LNCaP cell line showing the cell-cycle phase of LNCaP cell line for each cell.
(E) Histogram of LNCaP cell line showing the percentage of cell-cycle phase in each cluster.
(F) UMAP representation of the LNCaP cell line showing the contribution of siScrambled and siU6atac samples (on the left); the cell-cycle phase of siScrambled and siU6atac samples (on the center) and cell phase of each cell (on the right).
(G) UMAP representation of PM154 cell line showing the contribution of siScrambled and siU6atac samples (on the left); the cell-cycle phase of siScrambled and siU6atac samples (on the center) and cell phase of each cell.
(H) Histogram of LNCaP cell line showing the percentage of the cell-cycle phase for siScrambled and siU6atac cells. The table is showing the number of cells for each cell-cycle phase. The p value was calculated using a Fisher's exact test.
(I) Histogram of PM154 cell line showing the percentage of the cell-cycle phase for siScrambled and siU6atac cells. The table is showing the number of cells for each cell-cycle phase. The p value was calculated using a Fisher's exact test.
(J) UMAP representation of LNCaP cell line showing the AR score calculated on siScrambled and siU6atac samples. Violin plots show the AR score calculated for each cell-cycle phase. The p value was calculated using a Wilcoxon test.
(K) UMAP of LNCaP cell line showing the EMT score calculated on siScrambled and siU6atac samples. Violin plots show the EMT score calculated for each cell-cycle phase. The p value was calculated using a Wilcoxon test.
See also Figures S13–S15 and Table S13.
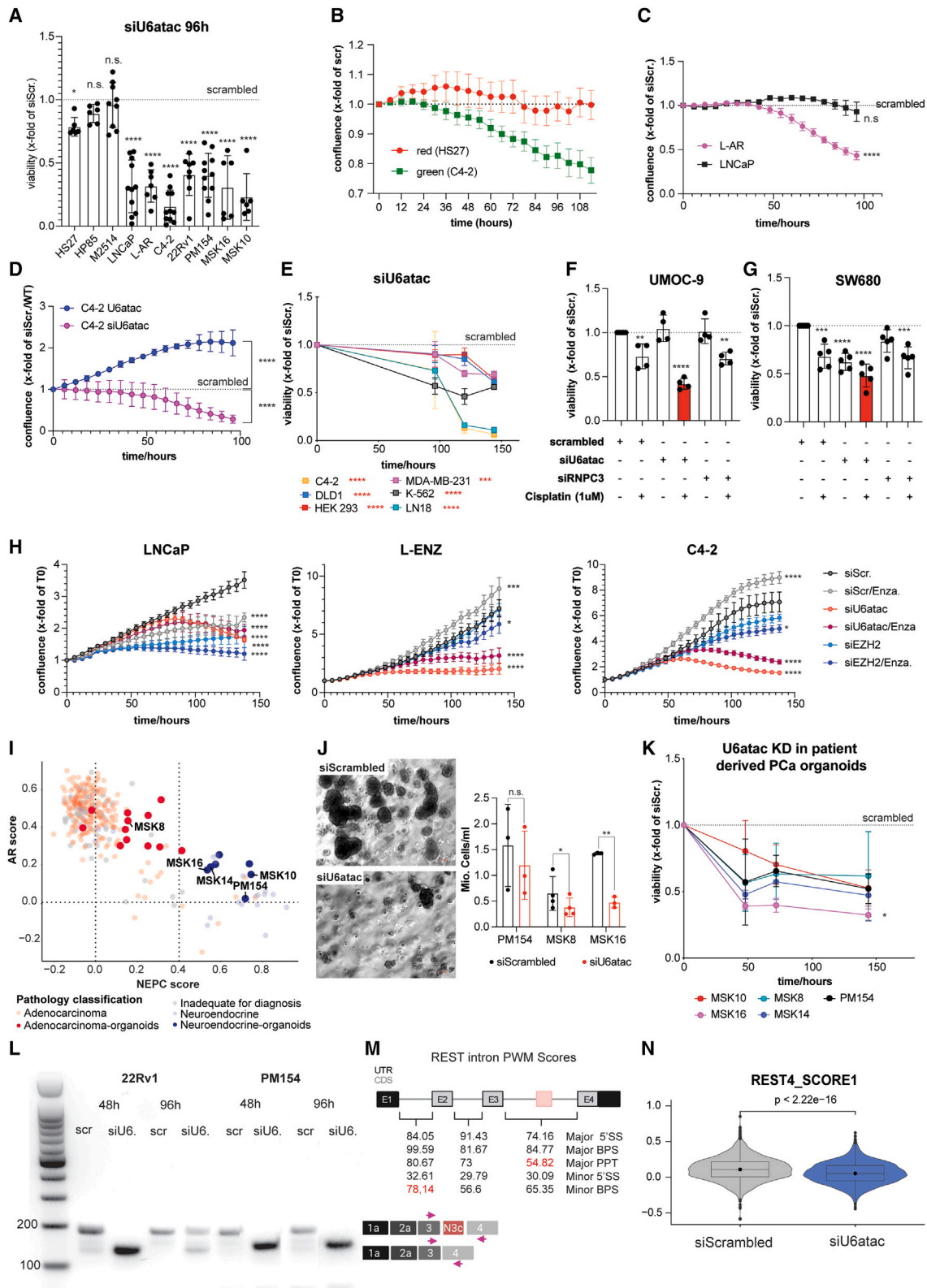
(legend on next page)

the EMT score was also significantly reduced in G1 (Wilcoxon test, p < 2.22e-16), G2M (Wilcoxon test, LNCaP: p = 2.9e-10, PM154: p = 2.9e-06), and S-phase (Wilcoxon test, LNCaP: p = 3.4e−07, PM154: p = 0.0089) in siU6atac-treated cells and organoids (Figures 5K and S13I). Next, we looked at the expression of MIGs identified as the most crucial nodes from our PPI analysis in the scRNA-seq data (Figures 1B and 1C). We found a significantly lower number of siU6atac-treated cells and organoids that expressed these MIGs, such as EED, ACTL6A, and PARP1 (Figures S14 and S15). Together, scRNA-seq revealed siU6atac-mediated transcriptomic remodeling that reduces PCa progression and lineage plasticity by decreasing cell cycle, AR signaling, and EMT.

The minor spliceosome is essential for PCa growth and viability

Although siU6atac-mediated inhibition of the MiS resulted in similar splice pattern shifts in normal prostate (HP85) cells, fibroblasts (HS27), and PCa cells (C4-2) (Figure S16), siU6atac only significantly decreased proliferation in PCa cells and organoids after 96 h (Figures 6A and S17A–S17C). These findings further underscore the context-dependent effects of MiS inhibition and the predicted cellular outcomes of shifts in AS events. To address tumor selectivity, we performed experiments where C4-2-GFP cancer cells and HS27-mCherry fibroblasts (Figures 6B and S17D) were co-cultured and subjected to siU6atac treatment. siU6atac decreased the green fluorescent signal

(C4-2) significantly but did not affect the red fluorescent signal (HS27) (Figures 6B, S17E, and S17F). We next compared the effect of siU6atac treatment on AR-sensitive and resistant cancer cells. We observed a significant decrease in the confluence of LNCaP (L) expressing AR (L-AR) cells compared with LNCaP cells upon siU6atac treatment (Figures 6C and S18A). L-AR cells showed a higher proliferation defect than LNCaP cells (Figure S17B). L-AR cells (as well as therapy-resistant cells) were shown to possess higher MiS activity than their native hormone-responsive lines (Figure 2). These data suggest that a higher basal MiS activity indicates an increased dependency on this non-canonical splicing mechanism. Overexpression of U6atac in C4-2 cells enables cell growth (Figure 6D), which agrees with previous findings showing that the MiS is essential for cancer cell proliferation.[51,52]

MiS inhibition also affects the viability of colon, kidney, breast, glioblastoma, and CML cell lines (Figures 6E, S18B, and S18C), suggesting that the essential role of MiS not restricted to PCa. Among all tested siRNAs (RNPC3, PDCD7, U12, and U6atac), U6atac KD was the most efficient. Even the KD of U6atac was insufficient to decrease the viability of the bladder cancer lines UMOC9 and SW680. However, siU6atac (in combination with cisplatin) led to a significant decrease in viability (Figures 6F and 6G). Based on pre-clinical reports that CRPC benefits from combination therapy (e.g., iEZH2i and the ARSi enzalutamide),[53,54] we tested U6atac inhibition plus enzalutamide. We observed that U6atac KD alone and in combination with

Figure 6. U6atac and the MiS represent potential therapeutic targets in cancer

(A) Cell viability in human normal (HP85 n = 6) and PCa cell lines (LNCaP n = 11, L-AR n = 7, C4-2 n = 11, 22Rv1 n = 8, PM154 n = 11, MSK16 n = 6, MSK10 n = 6), human fibroblast cells (HS27 n = 7), and primary mouse prostate cells (MS2514 n = 9) (mean ± SEM, ordinary two-way ANOVA; ns p > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001). Experiments were performed in triplicates.

(B) Growth curves of mCherry positive HS27 cells and GFP positive C4-2 cells grown in a co-culture, which was treated with siScrambled or siU6atac. Data represented as x fold of siScrambled. Data represent pooled results from biologically independent experiments (n = 3, mean ± SEM, ordinary two-way ANOVA; ns p > 0.05, *p < 0.05).

(C) Growth curves of LNCaP and L-AR cells treated for several time points with siU6atac. Data are normalized to the siScrambled control and represents pooled results from 4 biologically independent experiments (mean ± SEM, ordinary two-way ANOVA; ns p > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001).

(D) Growth curves of C4-2 cells stable overexpressing U6atac normalized to C4-2 cells expressing the EV plasmid or C4-2 cells treated for several time points with siU6atac normalized to C4-2 cells treated with siScrambled. Data represent pooled results from 16 (C4-2 siU6atac) and 4 (C4-2 U6atac) biologically independent experiments (mean ± SD ordinary two-way ANOVA; ns p > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001).

(E) Cell viability in different human cancer cell lines treated for several time points with siU6atac. Data are normalized to the siScrambled control and represent pooled results from 4 biologically independent experiments (mean ± SEM ordinary two-way ANOVA; ns p > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001).

(F) Cell viability in human bladder cancer UMOC-9 cells treated with siRNA against U6atac, RNPC3, and Scrambled or with cisplatin (1 μM). Data represent pooled results from 4 biologically independent experiments (mean ± SD, ordinary one-way ANOVA; ns p > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001).

(G) Cell viability in human bladder cancer SW680 cells treated with siRNA against U6atac, RNPC3, and Scrambled or with cisplatin (1 μM). Data represent pooled results from 4 biologically independent experiments (mean ± SD, ordinary one-way ANOVA; ***p < 0.001, ****p < 0.001).

(H) Growth curves of LNCaP (n = 4), L-ENZ (n = 6) and C4-2 (n = 4) cells treated with siScrambled, siU6atac and siEZH2 ± enzalutamide (10 μM). Data represent pooled results from biologically independent experiments (mean ± SEM, ordinary two-way ANOVA; ns p > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001).

(I) Scatterplot representing RNA-seq results from patient-derived PCa organoids (big dots) superimposing clinical data from the SU2C dataset (small dots). y axis represents the AR score, x axis represents the NEPC score.

(J) Bright-field microcopy of MSK8 cells 8 days after siRNA treatment against U6atac. Cells treated with Scrambled siRNA are shown as control. Scale bars (red): 50 μm. Bar blot summarizes cell counts at day 8 (n = 3, paired t test, *p = 0.0307).

(K) Cell viability in patient-derived PCa organoids. Data are normalized to the scrambled control and represent pooled results from three biologically independent experiments (mean ± SEM, two-way ANOVA; ns p > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001).

(L) Product of a qRT-PCR with primers binding to exon 3 (fw) and exon 4 (rv) of REST loaded on a 3% agarose gel. Experiment performed in 22Rv1 and PM154 cells and organoids treated with siScrambled (scr.) or siU6atac (48 and 96 h). canonical REST size expected at 150 bp and REST4 sample size expected at 200 bp. Image shows a composite in which MSK16 samples after 22Rv1 cell samples were cropped out. In addition, PM154 scrambled and siU6atac samples originally were positioned side by side and are now cropped together to match the 22Rv1 composition. Original image published through Mendeley data.

(M) PMW score calculations of minor and major intron consensus sequences in the REST gene. Abbreviations: 5′ SS, 5′ splice site; 3′ SS, 3′ splice site; PPT, poly pyrimidine tract BPS, branchpoint sequence.

(N) Violin plot shows the REST4 score calculated on siScrambled and siU6atac PM154 samples. The p value was calculated using a Wilcoxon test.

See also Figures S16–S19.

enzalutamide was more effective than EZH2 plus enzalutamide (Figures 6H and S18D). In conclusion, siU6atac is sufficient to block cancer cell survival.

Knocking down U6atac in 3D patient-derived organoids that reflect the range of advanced PCa (Figure 6I) showed similar results as with the 2D models. Treatment of the organoids MSK8,10,16,14, and PM154 with siU6atac provoked a significant reduction in organoid growth and viability (Figures 6J, 6K, and S18E). Collectively, these data show that knocking down MiS components is sufficient to decrease cancer growth and viability.

Previous reports demonstrate that 10%–15% of CRPC-adeno undergo lineage plasticity to an AR-negative state, often with NE features.[21,55] The REST is a repressor of neural differentiation. Loss of REST expression is associated with the upregulation of genes that define CRPC-NE.[56,57] Loss of REST expression precedes NE differentiation in PCa. REST is regulated through dynamic AS such that a microexon in intron 3, when included, results in a premature stop codon truncating the final protein product (Figure 6M). This truncated form of REST protein, referred to as REST4, cannot bind to the RE1 silencing element upstream of target genes but can block REST/DNA contact.[58] REST4 thus acts as a dominant negative, and consequently, increased levels of REST4 would inhibit REST function, enabling NE differentiation. We observed that REST expression by qPCR is dynamic across the four different cell lines representing PCa progression (LNCaP>C4-2>22Rv1>PM154) (Figure S19A, upper panel). When we compared REST4 expression, it was significantly upregulated in 22Rv1 and PM154 (Figure S19A, lower panel). Analysis of the SU2C data matches these results such that canonical REST positively correlates with the AR score and negatively with the NEPC score (Figures S19B and S19C). In agreement, REST expression was significantly decreased in CRPC-NE samples (NEPC score > 0.4) when compared with AR-positive samples (AR score > 0.4) (Figures S19B and S19C). By contrast, when we queried SU2C for REST isoform expression, we found that REST4 negatively correlates with AR and positively correlates with NEPC score. Here, REST4 expression was significantly increased in NEPC samples (NEPC score > 0.4) when compared with AR samples (AR score > 0.4) (Figures S19B and S19C).

When we KD U6atac, canonical REST increases and REST4 decreases expression in NE-like cells (22Rv1) and the NE-organoids (PM154) (Figures S19D–S19H). We performed RT-PCR with primers positioned in exons 3 and 4 and found that MiS inhibition through siU6atac resulted in the increase of canonical REST isoform (Figure 6L, lower band) and the downregulation of REST4 (Figure 6L, upper band) in 22Rv1 and PM154. These data were further confirmed by isoform-specific qRT-PCR analysis (Figures S19F and S19H).

The REST/REST4 switch on siU6atac-mediated MiS inhibition was surprising as REST is not characterized as a MIG. We discovered that the intron containing the N3c microexon has a reduced score for major poly pyrimidine tract (PPT) and a stronger score for the minor BPS (Figure 6M). Although this deviation does not make REST a MIG, it allows for a potential MiS involvement in regulating the AS of N3c microexon.

As described above, REST4 upregulation is a dominant-negative blocking REST function, which is expected to result in the downregulation of NE genes. Indeed, overexpression of REST4

in a non-NE setting (C4-2 cells) confirmed this. Being non-NE, C4-2 cells do not change REST/REST4 splicing upon U6atac manipulation in our data. Yet, when we overexpress REST4 in those cells, we observe a decrease in the canonical REST variant and an increase in SYP, CHGA, and SNAP25 (NE genes) (Figure S19G). Using Labrecque et al. REST4 score[59] on our PM154 scRNA-seq data further confirmed the decrease in REST4 score upon siU6atac treatment (Figure 6N). Taken together, an increase in REST by siU6atac should decrease the expression of NE genes. Indeed, qRT-PCR analysis showed that upon siU6atac treatment, the expression of NE genes, such as SYP, CHGA, VGF, and SNAP25, were downregulated in 22Rv1 and PM154 (Figure S19H). This finding provides insight into how MiS may enable the CRPC-adeno to CRPC-NE transition via the regulation of AS of REST.

DISCUSSION

MIG and MiS component expression discriminate cancers

MIG expression discriminates different cancer types, and in our study, we show the example of PCa progression. U6atac is tightly regulated by normal cells, and upregulation occurs in stress, but it also occurs, as seen in this study, in cancer. U6atac enhances MiS activity to control MIG expression "on demand," thereby underscoring the importance of U6atac and MiS function during cancer progression. Future work needs to explore how U6atac expression affects cancers from different origins differently.

MiS activity increases with cancer progression

Minor and not major intron-splicing efficiency varied across different PCa states, especially prolonged ARSi treatment of PCa. Although manipulation of AR signaling altered MiS activity as reported by the minor intron-splicing reporter, it is unlikely to be involved in CRPC-NE, which often exhibits low AR activity. Thus, we speculate that p38MAPK, a known regulator of U6atac turnover, might affect U6atac levels in CRPC-NE. Genes encoding p38MAPK family members are MIGs, thereby creating a feedback loop in which MiS activity is regulated by p38MAPK, which, in turn, is regulated by MiS. Similar regulatory loop between AR activity and MiS function through siU6atac was observed. Together, the regulation of MiS activity is engaged with PCa-relevant oncogenic pathways.

siU6atac impairs minor intron splicing

The large number of MIGs with high minor intron retention underscores the importance of U6atac in MiS function.[60,61] Increased cryptic splicing by the major spliceosome in siU6atac-treated samples suggests that 5′ SS recognition by U11/U12 di-snRNP is unaffected and that U4atac-U6atac-U5 tri-snRNP, which is recruited later, is compromised.[62,63] The significantly elevated intron retention in major introns flanking minor introns upon MiS inhibition suggests that minor and major spliceosome not only interact to establish the exon bridge[48] but to also complete the splicing reaction.

Minor intron retention and aberrant cryptic splicing were anticipated to be cell type specific and were predicted to compromise a myriad of MIG-regulated biological pathways, including

chromatin modification, transcription, splicing, and others.[19] Indeed, non-MIGs, such as AR, EZH2, and AURKA, were downregulated. By contrast, aberrantly spliced MIG transcripts with premature stop codons were not downregulated. We suspect NMD was itself compromised by aberrant splicing of UPF1 and NCBP2, two crucial NMD factors. Thus, aberrantly spliced MIG transcripts, which we have shown are bound to polysomes,[48] might produce truncated MIG proteins.

siU6atac triggers a context-dependent decrease in the expression of cancer-relevant MIGs and non-MIGs

The aberrant splicing of MIGs converges on the downregulation of non-MIGs that together enrich for cell-cycle regulation, which was not the case when MiS was inhibited in normal human fibroblasts or prostate cells. This finding supports our central thesis that the MiS is a potential therapeutic target for cancer.

MiS inhibition blocks the proliferation and survival of different PCa cell lines and organoids through the downregulation of MIGs in distinct biological pathways. Thus, MiS inhibition in different systems still ends up converging on the same biologically relevant endpoint for cancer (i.e., cell cycle and survival). Bolstering this idea is that MiS inhibition results in elevated AURKA (non-MIG) protein, which is important in CRPC differentiation and aggressiveness,[21,64] in therapy-sensitive LNCaP cells. However, it is decreased in CRPC- adeno/NE intermediate cells 22Rv1 and CRPC-NE organoid PM154. Similarly, the polycomb group protein EED (MIG), known to regulate AR expression levels[65] and a potential target in CRPC,[66,67] was decreased in C4-2 and 22Rv1 cells. By contrast, we observed increased protein levels for the G1 cell-cycle arrest mediator CDN1A (non-MIG) in all CRPC-adeno cells. We found that genes with downregulation of both transcript and protein are highly enriched for cell-cycle regulation and DDR GO terms. Thus, despite the dynamic fluctuations in mRNA and protein levels, the core molecular defect is the inhibition of proliferation. We also observed the downregulation of genes relevant to EMT, a hallmark of metastasis, after MiS inhibition. MiS inhibition also reduced AR activity in PCa cells and organoids. This supports our finding of a feedback loop between AR and MiS activity.

The minor spliceosome is essential for PCa growth and viability

Whereas siU6atac-mediated MiS inhibition substantially affected cancer cells, it did not strongly affect normal human prostate organoids and fibroblasts or benign mouse prostate cells. This finding has therapeutic implications because MiS inhibition is specifically detrimental to cancer cells and organoids. MiS dependency does not merely track the proliferative rate of cells. Instead, MiS dependency is cancer type and stage specific. Indeed, we also show that PCa-specific oncogenic drivers, such as AR, REST, and P38MAPK, are all intertwined with MiS activity and might play a role in defining MiS dependency. We uncover a potential insight into lineage plasticity, an important mechanism of PCa resistance. The effect of MiS inhibition on CRPC-NE was discovered to go through the aberrant splicing of REST, which is critical for CRPC-adeno to CRPC-NE transition. REST suppresses the expression of neuronal genes. We found that siU6atac treatment shifted the splice patterns of REST such that there was an increase in the ca-

nonical REST1 isoform, whereas REST4, which lacks the DNA-binding domain and is known to act as a dominant negative, was decreased. The outcome of REST/REST4 splicing shift is the restoration of REST function such that the targets of REST should be suppressed. Indeed, crucial NE genes were downregulated in siU6atac-treated NE models.

We observed that AR overexpression increased the proliferation of LNCaP cells, thereby making them more susceptible to siU6atac-mediated MiS inhibition. We posit that U6atac is a promising target for CRPC-adeno PCa and CRPC-NE, through regulation of AR and MAPK14 (p38MAPK), respectively, both major PCa-relevant pathways.[34,35,68,69] We explored some common treatment approaches for advanced PCa. We found that U6atac KD alone is sufficient to block the proliferation of CRPC-adeno cells, which is better than enzalutamide, siEZH2, or siEZH2/enzalutamide. We show that MiS inhibition also works in PCa patient-derived organoids (CRPC-adeno and CRPC-NE), whose transcriptome signature correlates with patients tested in the SU2C study.

We propose that our findings are not limited to PCa. Current work is now exploring how modulation of MIS can be effectively used to target other cancer. The strategies will be tissue context specific. For example, U6atac KD combined with cisplatin was effective in killing bladder cancer cells. Therefore, we anticipate that the successful use of MIS modulation will require exploring the greatest vulnerability for each cancer.

Overall, we show that MiS activity plays a crucial role in the progression and transformation of PCa and that inhibiting different MiS components can block cancer cell proliferation and viability, with U6atac as the most effective target. We show that siU6atac can successfully inhibit proliferation and viability through disruption of pathways such as MAPK, cell cycle, and DNA repair. Finally, we posit that MiS inhibition is a potential therapeutic target that extends beyond PCa to other cancers.

Limitations of the study

Although we successfully show that MiS inhibition through siU6atac successfully targets tumor cells while sparing normal cells, we recognize that this study lacks *in vivo* experimentation. We are currently developing modalities to deliver siU6atac to *in vivo* tumor models.

Given the numerous targets affected by siU6atac, it should also be noted that it is unlikely that the phenotype observed is the result of perturbations of a single MIG but rather a constellation of gene networks. Finally, although the minor/major spliceosome reporter assays lacked internal controls, we found a robust response without large fluctuations in our technical and biological replicates, indicating that the observed effects were not due to variations in transfections but due to biological differences. Moreover, the same approach for major intron-splicing reporter did not show the response specific to minor intron-splicing reporter.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2023.05.017.

REFERENCES

1. Argente, J., Abusrewil, S.A., Bona, G., Chiarelli, F., Kelnar, C.J., and Skordis, N.; International Workshop on Management of Puberty for Optimum Auxological Results (2001). Isolated growth hormone deficiency in children and adolescents. J. Pediatr. Endocrinol. Metab. *14* (*Suppl 2*), 1003–1008. https://doi.org/10.1515/jpem-2001-s213.

2. Gucev, Z., Tasic, V., Saranac, L., Stobbe, H., Kratzsch, J., Klammt, J., and Pfäffle, R. (2012). A novel GH1 mutation in a family with isolated growth hormone deficiency type II. Horm. Res. Paediatr. *77*, 200–204. https://doi.org/10.1159/000334643.

3. Elsaid, M.F., Chalhoub, N., Ben-Omran, T., Kumar, P., Kamel, H., Ibrahim, K., Mohamoud, Y., Al-Dous, E., Al-Azwani, I., Malek, J.A., et al. (2017). Mutation in noncoding RNA RNU12 causes early onset cerebellar ataxia. Ann. Neurol. *81*, 68–78. https://doi.org/10.1002/ana.24826.

4. Roifman, C.M., and Melamed, I. (2003). A novel syndrome of combined immunodeficiency, autoimmunity and spondylometaphyseal dysplasia. Clin. Genet. *63*, 522–529. https://doi.org/10.1034/j.1399-0004.2003.00033.x.

5. Edery, P., Marcaillou, C., Sahbatou, M., Labalme, A., Chastang, J., Touraine, R., Tubacher, E., Senni, F., Bober, M.B., Nampoothiri, S., et al. (2011). Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. Science *332*, 240–243. https://doi.org/10.1126/science.1202205.

6. He, H., Liyanarachchi, S., Akagi, K., Nagy, R., Li, J., Dietrich, R.C., Li, W., Sebastian, N., Wen, B., Xin, B., et al. (2011). Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. Science *332*, 238–240. https://doi.org/10.1126/science.1200587.

7. Resta, N., Pierannunzio, D., Lenato, G.M., Stella, A., Capocaccia, R., Bagnulo, R., Lastella, P., Susca, F.C., Bozzao, C., Loconte, D.C., et al. (2013). Cancer risk associated with STK11/LKB1 germline mutations in Peutz-Jeghers syndrome patients: results of an Italian multicenter study. Dig. Liver Dis. *45*, 606–611. https://doi.org/10.1016/j.dld.2012.12.018.

8. Fischer, D., Wahlfors, T., Mattila, H., Oja, H., Tammela, T.L., and Schleutker, J. (2015). MiRNA profiles in lymphoblastoid cell lines of Finnish prostate cancer families. PLoS One *10*. e0127427. https://doi.org/10.1371/journal.pone.0127427.

9. Takayama, K.I., Suzuki, T., Fujimura, T., Yamada, Y., Takahashi, S., Homma, Y., Suzuki, Y., and Inoue, S. (2017). Dysregulation of spliceosome gene expression in advanced prostate cancer by RNA-binding

protein PSF. Proc. Natl. Acad. Sci. USA *114*, 10461–10466. https://doi.org/10.1073/pnas.1706076114.

10. Verma, B., Akinyi, M.V., Norppa, A.J., and Frilander, M.J. (2018). Minor spliceosome and disease. Semin. Cell Dev. Biol. *79*, 103–112. https://doi.org/10.1016/j.semcdb.2017.09.036.

11. Chandrasekar, T., Yang, J.C., Gao, A.C., and Evans, C.P. (2015). Mechanisms of resistance in castration-resistant prostate cancer (CRPC). Transl. Androl. Urol. *4*, 365–380. https://doi.org/10.3978/j.issn.2223-4683.2015.05.02.

12. Vlachostergios, P.J., Puca, L., and Beltran, H. (2017). Emerging variants of castration-resistant prostate cancer. Curr. Oncol. Rep. *19*, 32. https://doi.org/10.1007/s11912-017-0593-6.

13. Clermont, P.-L., Ci, X., Pandha, H., Wang, Y., and Crea, F. (2019). Treatment-emergent neuroendocrine prostate cancer: molecularly driven clinical guidelines. Int. J. Endocr. Oncol. *6*. https://doi.org/10.2217/ije-2019-0008.

14. Lee, A.R., Che, N., Lovnicki, J.M., and Dong, X. (2018). Development of neuroendocrine prostate cancers by the ser/Arg repetitive matrix 4-mediated RNA splicing network. Front. Oncol. *8*, 93. https://doi.org/10.3389/fonc.2018.00093.

15. Zhang, D., Hu, Q., Liu, X., Ji, Y., Chao, H.P., Liu, Y., Tracz, A., Kirk, J., Buonamici, S., Zhu, P., et al. (2020). Intron retention is a hallmark and spliceosome represents a therapeutic vulnerability in aggressive prostate cancer. Nat. Commun. *11*, 2089. https://doi.org/10.1038/s41467-020-15815-7.

16. Giridhar, K., Sosa, C., Sicotte, H., Wang, L., Wang, L., Sinnwell, J.P., Tan, W., Costello, B.A., Quevedo, F., Pitot, H.C., et al. (2019). Evolution of androgen receptor variant (ARV) profiles in serial metastatic solid and liquid biopsies in metastatic castrate resistant prostate cancer (mCRPC) treated with abiraterone acetate/ prednisone (AA/P). J. Clin. Oncol. *37*. e16559. https://doi.org/10.1200/JCO.2019.37.15_suppl.e16559.

17. Alioto, T.S. (2007). U12DB: a database of orthologous U12-type spliceosomal introns. Nucleic Acids Res. *35*. D110–D115. https://doi.org/10.1093/nar/gkl796.

18. Baumgartner, M., Drake, K., and Kanadia, R.N. (2019). An integrated model of minor intron emergence and conservation. Front. Genet. *10*, 1113. https://doi.org/10.3389/fgene.2019.01113.

19. Olthof, A.M., Hyatt, K.C., and Kanadia, R.N. (2019). Minor intron splicing revisited: identification of new minor intron-containing genes and tissue-dependent retention and alternative splicing of minor introns. BMC Genomics *20*, 686. https://doi.org/10.1186/s12864-019-6046-x.

20. Tamborero, D., Rubio-Perez, C., Deu-Pons, J., Schroeder, M.P., Vivancos, A., Rovira, A., Tusquets, I., Albanell, J., Rodon, J., Tabernero, J., et al. (2018). Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med. *10*, 25. https://doi.org/10.1186/s13073-018-0531-8.

21. Beltran, H., Prandi, D., Mosquera, J.M., Benelli, M., Puca, L., Cyrta, J., Marotz, C., Giannopoulou, E., Chakravarthi, B.V., Varambally, S., et al. (2016). Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nat. Med. *22*, 298–305. https://doi.org/10.1038/nm.4045.

22. Belorkar, A., and Wong, L. (2016). GFS: fuzzy preprocessing for effective gene expression analysis. BMC Bioinformatics *17* (*Suppl 17*), 540. https://doi.org/10.1186/s12859-016-1327-8.

23. Van Laere, S.J., Ueno, N.T., Finetti, P., Vermeulen, P., Lucci, A., Robertson, F.M., Marsan, M., Iwamoto, T., Krishnamurthy, S., Masuda, H., et al. (2013). Uncovering the molecular secrets of inflammatory breast cancer biology: an integrated analysis of three distinct Affymetrix gene expression datasets. Clin. Cancer Res. *19*, 4685–4696. https://doi.org/10.1158/1078-0432.CCR-12-2549.

24. Zhao, S., Sun, J., Shimizu, K., and Kadota, K. (2018). Silhouette scores for arbitrary defined groups in gene expression data and insights into differential expression results. Biol. Proced. Online *20*, 5. https://doi.org/10.1186/s12575-018-0067-8.

25. Lovmar, L., Ahlford, A., Jonsson, M., and Syvänen, A.C. (2005). Silhouette scores for assessment of SNP genotype clusters. BMC Genomics *6*, 35. https://doi.org/10.1186/1471-2164-6-35.

26. McMahan, Z.H., Domsic, R.T., Zhu, L., Medsger, T.A., Casciola-Rosen, L., and Shah, A.A. (2019). Anti-RNPC-3 (U11/U12) antibodies in systemic sclerosis in patients with moderate-to-severe gastrointestinal dysmotility. Arthritis Care Res. (Hoboken) *71*, 1164–1170. https://doi.org/10.1002/acr.23763.

27. Shah, A.A., Xu, G., Rosen, A., Hummers, L.K., Wigley, F.M., Elledge, S.J., and Casciola-Rosen, L. (2017). Brief report: anti–RNPC-3 antibodies as a marker of cancer-associated scleroderma. Arthritis Rheumatol. *69*, 1306–1312. https://doi.org/10.1002/art.40065.

28. Younis, I., Berg, M., Kaida, D., Dittmar, K., Wang, C., and Dreyfuss, G. (2010). Rapid-response splicing reporter screens identify differential regulators of constitutive and alternative splicing. Mol. Cell. Biol. *30*, 1718–1728. https://doi.org/10.1128/MCB.01301-09.

29. Puca, L., Bareja, R., Prandi, D., Shaw, R., Benelli, M., Karthaus, W.R., Hess, J., Sigouros, M., Donoghue, A., Kossai, M., et al. (2018). Patient derived organoids to model rare prostate cancer phenotypes. Nat. Commun. *9*, 2404. https://doi.org/10.1038/s41467-018-04495-z.

30. Gao, D., Vela, I., Sboner, A., Iaquinta, P.J., Karthaus, W.R., Gopalan, A., Dowling, C., Wanjala, J.N., Undvall, E.A., Arora, V.K., et al. (2014). Organoid cultures derived from patients with advanced prostate cancer. Cell *159*, 176–187. https://doi.org/10.1016/j.cell.2014.08.016.

31. Mertz, K.D., Setlur, S.R., Dhanasekaran, S.M., Demichelis, F., Perner, S., Tomlins, S., Tchinda, J., Laxman, B., Vessella, R.L., Beroukhim, R., et al. (2007). Molecular characterization of TMPRSS2-ERG gene fusion in the NCI-H660 prostate cancer cell line: a new perspective for an old model. Neoplasia *9*, 200–206. https://doi.org/10.1593/neo.07103.

32. Qin, S., Liu, D., Kohli, M., Wang, L., Vedell, P.T., Hillman, D.W., Niu, N., Yu, J., Weinshilboum, R.M., and Wang, L. (2018). TSPYL family regulates CYP17A1 and CYP3A4 expression: potential mechanism contributing to abiraterone response in metastatic castration-resistant prostate cancer. Clin. Pharmacol. Ther. *104*, 201–210. https://doi.org/10.1002/cpt.907.

33. Younis, I., Dittmar, K., Wang, W., Foley, S.W., Berg, M.G., Hu, K.Y., Wei, Z., Wan, L., and Dreyfuss, G. (2013). Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. eLife *2*. e00780. https://doi.org/10.7554/eLife.00780.

34. Frank, S.B., Berger, P.L., Ljungman, M., and Miranti, C.K. (2017). Human prostate luminal cell differentiation requires notch3 induction by p38-MAPK and MYC. J. Cell Sci. *130*, 1952–1964. https://doi.org/10.1242/jcs.197152.

35. Cuenda, A., and Sanz-Ezquerro, J.J. (2017). p38γ and p38δ: from spectators to key physiological players. Trends Biochem. Sci. *42*, 431–442. https://doi.org/10.1016/j.tibs.2017.02.008.

36. Abad, X., Razquin, N., Abad, A., and Fortes, P. (2010). Combination of RNA interference and U1 inhibition leads to increased inhibition of gene expression. Nucleic Acids Res. *38*. e136. https://doi.org/10.1093/nar/gkq299.

37. Berezhna, S.Y., Supekova, L., Supek, F., Schultz, P.G., and Deniz, A.A. (2006). siRNA in human cells selectively localizes to target RNA sites. Proc. Natl. Acad. Sci. USA *103*, 7682–7687. https://doi.org/10.1073/pnas.0600148103.

38. Avivi, S., Mor, A., Dotan, I., Tzadok, S., Kanter, I., Kinor, N., Canaani, D., and Shav-Tal, Y. (2017). Visualizing nuclear RNAi activity in single living human cells. Proc. Natl. Acad. Sci. USA *114*. E8837–E8846. https://doi.org/10.1073/pnas.1707440114.

39. Castel, S.E., and Martienssen, R.A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat. Rev. Genet. *14*, 100–112. https://doi.org/10.1038/nrg3355.

40. Gagnon, K.T., Li, L., Chu, Y., Janowski, B.A., and Corey, D.R. (2014). RNAi factors are present and active in human cell nuclei. Cell Rep. *6*, 211–221. https://doi.org/10.1016/j.celrep.2013.12.013.

41. Guang, S., Bochner, A.F., Pavelec, D.M., Burkhart, K.B., Harding, S., Lachowiec, J., and Kennedy, S. (2008). An Argonaute transports siRNAs from the cytoplasm to the nucleus. Science *321*, 537–541. https://doi.org/10.1126/science.1157647.

42. Hoffer, P., Ivashuta, S., Pontes, O., Vitins, A., Pikaard, C., Mroczka, A., Wagner, N., and Voelker, T. (2011). Posttranscriptional gene silencing in nuclei. Proc. Natl. Acad. Sci. USA *108*, 409–414. https://doi.org/10.1073/pnas.1009805108.

43. Ji, F., Liu, L., Tien, Y.H., Peng, Y.H., and Lee, H.C. (2015). Non-catalytic RISCs and kinetics determine mammalian siRNA sub-cellular localization. PLoS One *10*. e0143182. https://doi.org/10.1371/journal.pone.0143182.

44. Robb, G.B., Brown, K.M., Khurana, J., and Rana, T.M. (2005). Specific and potent RNAi in the nucleus of human cells. Nat. Struct. Mol. Biol. *12*, 133–137. https://doi.org/10.1038/nsmb886.

45. Yahara, M., Kitamura, A., and Kinjo, M. (2017). U6 snRNA expression prevents toxicity in TDP-43-knockdown cells. PLoS One *12*. e0187813. https://doi.org/10.1371/journal.pone.0187813.

46. Haaland, R.E., Herrmann, C.H., and Rice, A.P. (2005). siRNA depletion of 7SK snRNA induces apoptosis but does not affect expression of the HIV-1 LTR or P-TEFb-dependent cellular genes. J. Cell. Physiol. *205*, 463–470. https://doi.org/10.1002/jcp.20528.

47. Bandiera, R., Wagner, R.E., Britto-Borges, T., Dieterich, C., Dietmann, S., Bornelöv, S., and Frye, M. (2021). RN7SK small nuclear RNA controls bidirectional transcription of highly expressed gene pairs in skin. Nat. Commun. *12*, 5864. https://doi.org/10.1038/s41467-021-26083-4.

48. Olthof, A.M., White, A.K., Mieruszynski, S., Doggett, K., Lee, M.F., Chakroun, A., Abdel Aleem, A.K., Rousseau, J., Magnani, C., Roifman, C.M., et al. (2021). Disruption of exon-bridging interactions between the minor and major spliceosomes results in alternative splicing around minor introns. Nucleic Acids Res. *49*, 3524–3545. https://doi.org/10.1093/nar/gkab118.

49. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189–196. https://doi.org/10.1126/science.aad0501.

50. Dong, B., Miao, J., Wang, Y., Luo, W., Ji, Z., Lai, H., Zhang, M., Cheng, X., Wang, J., Fang, Y., et al. (2020). Single-cell analysis supports a luminal-neuroendocrine transdifferentiation in human prostate cancer. Commun. Biol. *3*, 778. https://doi.org/10.1038/s42003-020-01476-1.

51. Jung, H.J., and Kang, H. (2014). The Arabidopsis U11/U12-65K is an indispensible component of minor spliceosome and plays a crucial role in U12 intron splicing and plant development. Plant J. *78*, 799–810. https://doi.org/10.1111/tpj.12498.

52. König, H., Matter, N., Bader, R., Thiele, W., and Müller, F. (2007). Splicing segregation: the minor spliceosome acts outside the nucleus and controls cell proliferation. Cell *131*, 718–729. https://doi.org/10.1016/j.cell.2007.09.043.

53. Ku, S.Y., Rosario, S., Wang, Y., Mu, P., Seshadri, M., Goodrich, Z.W., Goodrich, M.M., Labbé, D.P., Gomez, E.C., Wang, J., et al. (2017). Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. Science *355*, 78–83. https://doi.org/10.1126/science.aah4199.

54. Bai, Y., Zhang, Z., Cheng, L., Wang, R., Chen, X., Kong, Y., Feng, F., Ahmad, N., Li, L., and Liu, X. (2019). Inhibition of enhancer of zeste homolog 2 (EZH2) overcomes enzalutamide resistance in castration-resistant prostate cancer. J. Biol. Chem. *294*, 9911–9923. https://doi.org/10.1074/jbc.RA119.008152.

55. Abida, W., Cyrta, J., Heller, G., Prandi, D., Armenia, J., Coleman, I., Cieslik, M., Benelli, M., Robinson, D., Van Allen, E.M., et al. (2019). Genomic correlates of clinical outcome in advanced prostate cancer. Proc. Natl. Acad. Sci. USA *116*, 11428–11436. https://doi.org/10.1073/pnas.1902651116.

56. Hohl, M., and Thiel, G. (2005). Cell type-specific regulation of RE-1 silencing transcription factor (REST) target genes. Eur. J. Neurosci. *22*, 2216–2230. https://doi.org/10.1111/j.1460-9568.2005.04404.x.

57. Lietz, M., Hohl, M., and Thiel, G. (2003). RE-1 silencing transcription factor (REST) regulates human synaptophysin gene transcription through an intronic sequence-specific DNA-binding site. Eur. J. Biochem. *270*, 2–9. https://doi.org/10.1046/j.1432-1033.2003.03360.x.

58. Ohnishi, T., Shirane, M., and Nakayama, K.I. (2017). SRRM4-dependent neuron-specific alternative splicing of protrudin transcripts regulates neurite outgrowth. Sci. Rep. *7*, 41130. https://doi.org/10.1038/srep41130.

59. Labrecque, M.P., Brown, L.G., Coleman, I.M., Lakely, B., Brady, N.J., Lee, J.K., Nguyen, H.M., Li, D., Hanratty, B., Haffner, M.C., et al. (2021). RNA splicing factors SRRM3 and SRRM4 distinguish molecular phenotypes of castration-resistant neuroendocrine prostate cancer. Cancer Res. *81*, 4736–4750. https://doi.org/10.1158/0008-5472.CAN-21-0307.

60. Drake, K.D., Lemoine, C., Aquino, G.S., Vaeth, A.M., and Kanadia, R.N. (2020). Loss of U11 small nuclear RNA in the developing mouse limb results in micromelia. Development *147*. https://doi.org/10.1242/dev.190967.

61. White, A.K., Baumgartner, M., Lee, M.F., Drake, K.D., Aquino, G.S., and Kanadia, R.N. (2021). Trp53 ablation fails to prevent microcephaly in mouse pallium with impaired minor intron splicing. Development *148*. https://doi.org/10.1242/dev.199591.

62. Yu, Y.T., and Steitz, J.A. (1997). Site-specific crosslinking of mammalian U11 and u6atac to the 5′ splice site of an AT-AC intron. Proc. Natl. Acad. Sci. USA *94*, 6030–6035. https://doi.org/10.1073/pnas.94.12.6030.

63. Frilander, M.J., and Steitz, J.A. (2001). Dynamic exchanges of RNA interactions leading to catalytic core formation in the U12-dependent spliceosome. Mol. Cell *7*, 217–226. https://doi.org/10.1016/s1097-2765(01)00169-1.

64. Beltran, H., Rickman, D.S., Park, K., Chae, S.S., Sboner, A., MacDonald, T.Y., Wang, Y., Sheikh, K.L., Terry, S., Tagawa, S.T., et al. (2011). Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. Cancer Discov. *1*, 487–495. https://doi.org/10.1158/2159-8290.CD-11-0130.

65. Liu, Q., Wang, G., Li, Q., Jiang, W., Kim, J.S., Wang, R., Zhu, S., Wang, X., Yan, L., Yi, Y., et al. (2019). Polycomb group proteins EZH2 and EED directly regulate androgen receptor in advanced prostate cancer. Int. J. Cancer *145*, 415–426. https://doi.org/10.1002/ijc.32118.

66. Li, X., Gera, L., Zhang, S., Chen, Y., Lou, L., Wilson, L.M., Xie, Z.R., Sautto, G., Liu, D., Danaher, A., et al. (2021). Pharmacological inhibition of noncanonical EED-EZH2 signaling overcomes chemoresistance in prostate cancer. Theranostics *11*, 6873–6890. https://doi.org/10.7150/thno.49235.

67. Dong, H., Liu, S., Zhang, X., Chen, S., Kang, L., Chen, Y., Ma, S., Fu, X., Liu, Y., Zhang, H., et al. (2019). An allosteric PRC2 inhibitor targeting EED suppresses tumor progression by modulating the immune response. Cancer Res. *79*, 5587–5596. https://doi.org/10.1158/0008-5472.CAN-19-0428.

68. Cui, Y., Sun, Y., Hu, S., Luo, J., Li, L., Li, X., Yeh, S., Jin, J., and Chang, C. (2016). Neuroendocrine prostate cancer (NEPCa) increased the neighboring PCa chemoresistance via altering the PTHrP/p38/Hsp27/androgen receptor (AR)/p21 signals. Oncogene *35*, 6065–6076. https://doi.org/10.1038/onc.2016.135.

69. Ware, K.E., Gupta, S., Eng, J., Kemeny, G., Puviindran, B.J., Foo, W.-C., Crawford, L.A., Garland Almquist, R., Runyambo, D., Thomas, B.C., et al. (2020). Convergent evolution of p38/MAPK activation in hormone resistant prostate cancer mediates pro-survival, immune evasive, and metastatic phenotypes. Preprint at bioRxiv. https://doi.org/10.1101/2020.04.22.050385.

70. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, et al. (2017) Genetic effects on gene expression across human tissues. Nature 550, 204–213. https://doi.org/10.1038/nature24277.

71. Reber, S., Stettler, J., Filosa, G., Colombo, M., Jutzi, D., Lenzken, S.C., Schweingruber, C., Bruggmann, R., Bachi, A., Barabino, S.M., et al. (2016). Minor intron splicing is regulated by FUS and affected by ALS-associated FUS mutants. EMBO J. *35*, 1504–1521. https://doi.org/10.15252/embj.201593791.

72. Das, J., and Yu, H. (2012). HINT: high-quality protein interactomes and their applications in understanding human disease. BMC Syst. Biol. *6*, 92. https://doi.org/10.1186/1752-0509-6-92.

73. Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comp. Appl. Math. *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

74. Mandric, I., Temate-Tiagueu, Y., Shcheglova, T., Al Seesi, S., Zelikovsky, A., and Mandoiu, I.I. (2017). Fast bootstrapping-based estimation of confidence intervals of expression levels and differential expression from RNA-Seq data. Bioinformatics *33*, 3302–3304. https://doi.org/10.1093/bioinformatics/btx365.

75. Middleton, R., Gao, D., Thomas, A., Singh, B., Au, A., Wong, J.J.L., Bomane, A., Cosson, B., Eyras, E., Rasko, J.E.J., et al. (2017). IRFinder: assessing the impact of intron retention on mammalian gene expression. Genome Biol. *18*, 51. https://doi.org/10.1186/s13059-017-1184-4.

76. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452. https://doi.org/10.1093/nar/gku1003.

77. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527. https://doi.org/10.1038/nbt.3519.

78. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

79. Olthof, A.M., Schwoerer, C.F., Weber, A.L., Arokiadhas, I., Doggett, K., Mieruszynski, S., Cnaani, A., Heath, J.K., Biran, J., and Kanadia, R.N. (2022). Taxonomy of introns, their evolution, and the role of minor introns in stress response. Preprint at bioRxiv. https://doi.org/10.1101/2022.10.12.511939.

80. Sherman, B.T., Hao, M., Qiu, J., Jiao, X., Baseler, M.W., Lane, H.C., Imamichi, T., and Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res *50*, W216–W221. https://doi.org/10.1093/nar/gkac194.

81. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

82. Zappia, L., and Oshlack, A. (2018). Clustering trees: a visualization for evaluating clusterings at multiple resolutions. GigaScience *7*. https://doi.org/10.1093/gigascience/giy083.

83. Ross, A.E., Johnson, M.H., Yousefi, K., Davicioni, E., Netto, G.J., Marchionni, L., Fedor, H.L., Glavaris, S., Choeurng, V., Buerki, C., et al. (2016). Tissue-based genomics augments post-prostatectomy risk stratification in a natural history cohort of intermediate- and high-risk men. Eur. Urol. *69*, 157–165. https://doi.org/10.1016/j.eururo.2015.05.042.

84. Karthaus, W.R., Iaquinta, P.J., Drost, J., Gracanin, A., van Boxtel, R., Wongvipat, J., Dowling, C.M., Gao, D., Begthel, H., Sachs, N., et al. (2014). Identification of multipotent luminal progenitor cells in human prostate organoid cultures. Cell *159*, 163–175. https://doi.org/10.1016/j.cell.2014.08.017.

85. Rodriguez-Calero, A., Gallon, J., Akhoundova, D., Maletti, S., Ferguson, A., Cyrta, J., Amstutz, U., Garofoli, A., Paradiso, V., Tomlins, S.A., et al. (2022). Alterations in homologous recombination repair genes in prostate cancer brain metastases. Nat. Commun. *13*, 2400. https://doi.org/10.1038/s41467-022-30003-5.

86. Briganti, A., Spahn, M., Joniau, S., Gontero, P., Bianchi, M., Kneitz, B., Chun, F.K., Sun, M., Graefen, M., Abdollah, F., et al. (2013). Impact of age and comorbidities on long-term survival of patients with high-risk prostate cancer treated with radical prostatectomy: a multi-institutional competing-risks analysis. Eur. Urol. *63*, 693–701. https://doi.org/10.1016/j.eururo.2012.08.054.

87. Tosco, L., Laenen, A., Briganti, A., Gontero, P., Karnes, R.J., Bastian, P.J., Chlosta, P., Claessens, F., Chun, F.K., Everaerts, W., et al. (2018). The EMPaCT classifier: A validated tool to predict postoperative prostate cancer-related death using competing-risk analysis. Eur. Urol. Focus *4*, 369–375. https://doi.org/10.1016/j.euf.2016.12.008.

88. Chys, B., Devos, G., Everaerts, W., Albersen, M., Moris, L., Claessens, F., De Meerleer, G., Haustermans, K., Briganti, A., Chlosta, P., et al. (2020). Preoperative risk-stratification of high-risk prostate cancer: A multicenter analysis. Front. Oncol. *10*, 246. https://doi.org/10.3389/fonc.2020.00246.

89. Schindelin, J., Rueden, C.T., Hiner, M.C., and Eliceiri, K.W. (2015). The ImageJ ecosystem: an open platform for biomedical image analysis. Mol. Reprod. Dev. *82*, 518–529. https://doi.org/10.1002/mrd.22489.

90. Dawson, H., Christe, L., Eichmann, M., Reinhard, S., Zlobec, I., Blank, A., and Lugli, A. (2020). Tumour budding/T cell infiltrates in colorectal cancer: proposal of a novel combined score. Histopathology *76*, 572–580. https://doi.org/10.1111/his.14006.

91. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

92. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for the ENCODE Project. Genome Res. *22*, 1760–1774. https://doi.org/10.1101/gr.135350.111.

93. Chen, Y., Lun, A.T., and Smyth, G.K. (2016). From reads to genes to pathways: differential expression analysis of RNA-seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Res *5*, 1438. https://doi.org/10.12688/f1000research.8987.2.

94. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

95. Wicckham H. ggPlot2: Elegant graphics for data analysis (use R!). Springer. https://doi.org/10.1007/978-0-387-98141-3.

96. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

97. Bunis, D.G., Andrews, J., Fragiadakis, G.K., Burt, T.D., and Sirota, M. (2021). dittoSeq: universal user-friendly single-cell and bulk RNA sequencing visualization toolkit. Bioinformatics *36*, 5535–5536. https://doi.org/10.1093/bioinformatics/btaa1011.

98. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360. https://doi.org/10.1038/nmeth.3317.

99. Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: database for annotation, visualization, and integrated discovery. Genome Biol. *4*. P3.

100. Metsalu, T., and Vilo, J. (2015). ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. Nucleic Acids Res. *43*. W566–W570. https://doi.org/10.1093/nar/gkv468.

101. Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. *34*, 267–273. https://doi.org/10.1038/ng1180.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
See Table S14	This paper	N/A
Bacterial and virus strains		
One Shot Mach1 T1 Phage-Resistant Chemically Competent E. coli cells	Invitrogen	C862003
pLenti-GIII-CMV-GFP-2A-Puro	ABM	LV290591
pLV[Exp]-EGFP:T2A:Puro-CMV>63-REST4	Vector builder	N/A
Biological samples		
Tissue micro-arrays PCBM cohort Bern	GTEx Consortium et al.[70]	N/A
Tissue micro-arrays primary prostate	European Multicenter High Risk Prostate Cancer Clinical and Translational research group (EMPaCT)	N/A
Chemicals, peptides, and recombinant proteins		
FBS	Gibco	10270106
penicillin-streptomycin	Gibco	11548876
Insulin	Sigma-Aldrich	I9278
Apo-Transferrin	Sigma-Aldrich	T1147
Hydrocortisone	Sigma-Aldrich	H6909
beta-estradiol	Sigma-Aldrich	E2257
GlutaMAX	Thermo Fisher Scientific	35050061
HEPES	Thermo Fisher Scientific	15630056
AA	Life Technologies	15240-062
A-83-01	Tocris	29-391-0
SB202190	Sigma-Aldrich	S7076
Recombinant Murine EGF	PeproTech	315-09
Human Recombinant FGF-10	Peprotech	100–26
Nicotinaminde	Sigma-Aldrich	N0636
Testosterone	Fluka	10300
Noggin	PeproTech	120-10C
R-spondin	PeproTech	315-32
growth factor-reduced Matrigel	VWR	BDAA356239
Enzalutamide	Selleck Chemicals	S1250
Lipofectamine RNAiMAX Transfection Reagent	Thermo Fisher Scientific	13778150
Lipofectamine 3000 Transfection Reagent	Thermo Fisher Scientific	L3000015
Critical commercial assays		
ReliaPrep miRNA Cell and Tissue Miniprep System	Promega	Z6212
FIREScript RT cDNA Synthesis Kit	Solis BioDyne	06-15-00200
HOT FIREPol EvaGreen qPCR Mix Plus	Solis BioDyne	08-24-00020
Platinum SuperFi II Green PCR Master Mix	Thermo Scientific	12369010
Chromium Next GEM Single Cell 3′ Reagent Kit v3.1	10X Genomics	PN-1000121
Illumina Stranded Total RNA	Illumina	20040529
NovaSeq 6000 S2 Reagent Kit v1.5	Illumina	20028316
NovaSeq 6000 S4 Reagent Kit v1.5	Illumina	20028313
Click-iT™ EdU Alexa Fluor™ 488 Flow Cytometry Assay Kit	Thermo Fisher Scientific	C10420

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CellTiter-Glo Luminescent Cell Viability Assay	Promega	G9243
Dual-Glo® Luciferase Assay System	Promega	E2940
Deposited data		
Bulk RNAseq and scRNAseq data (European Genome-phenome Archive)	This paper	EGAS00001005546
Mass spectrometry proteomic data (ProteomXchange)	This paper	PXD026949
Unprocessed image	This paper	https://doi.org/10.17632/28j9zzdh3k.1
Experimental models: Cell and Organoid lines		
Human: PNT1A cells	Sigma Aldrich	CB_95012614
Human: LNCaP cells	ATCC	CVCL_1379
Human: C4-2 cells	ATCC	CRL-3314
Human: 22Rv1 cells	ATCC	CRL-2505
Human: PC3 cells	ATCC	CRL-3470
Human: DLD-1 cells	ATCC	CCL-221
Human: L-rENZ cells	This paper	N/A
Human: L-AR cells	Reber et al.[71]	N/A
Human: RWPE cells	ATCC	CVCL_3791
Human: HEK293T cells	ATCC	CVCL_0063
Human: VCaP cells	ATCC	CRL-2876
Human: MDA-MB-231 cells	ATCC	HTB-26
Human: K-562 cells	ATCC	CCL-243
Human: PC-3M-Pro4 cells	ATCC	CVCL_D579
Human: DU145 cells	ATCC	CVCL_0105
Human: NCI-H660 cells	ATCC	CRL-5813
Human: HP85 organoid	Das and Yu[72]	N/A
Human: MSKCC-PCa8 organoid	Rousseeuw[73]	N/A
Human: MSKCC-PCa10 organoid	Rousseeuw[73]	N/A
Human: MSKCC-PCa14 organoid	Rousseeuw[73]	N/A
Human: MSKCC-PCa16 organoid	Rousseeuw[73]	N/A
Oligonucleotides		
See Table S14	This paper	N/A
Recombinant DNA		
CMV-luc2CP/ARE reporter	Shah et al.[27]	N/A
CMV-luc2CP empty vector backbone	Shah et al.[27]	N/A
luc1CFH4 reporter	Shah et al.[27]	N/A
Software and algorithms		
ImageJ		https://imagej.nih.gov/ij/
FlowJo 10.7.1	BD Life Sciences	https://www.flowjo.com
Minor intron retention pipeline	Olthof et al.[48]	https://github.com/amolthof/minor-intron-retention
STAR	Mandric et al.[74]	https://github.com/alexdobin/STAR/releases
Gencode	Middleton et al.[75]	https://www.gencodegenes.org
Cell ranger analysis pipeline v6.0.1	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/installation
Seurat 4.0.3	Szklarczyk et al.[76]	https://github.com/satijalab/seurat
R	R Core Team	https://www.r-project.org
Hisat2	Bray et al.[77]	http://www.ccb.jhu.edu/software/hisat/index.shtml

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
IsoEM2 and IsoDE2	Love et al.[78]	https://github.com/mandricigor/isoem2 https://toolshed.g2.bx.psu.edu/view/saharlcc/isoem2_isode2/
IRFinder	Olthof et al. et al.[79]	https://github.com/williamritchie/IRFinder
DAVID	Sherman et al.[80]	https://david.ncifcrf.gov
STRING	N/A	https://string-db.org
GSEA v4.0.3	N/A	https://www.gsea-msigdb.org/gsea/index.jsp
MaxQuant 1.6.1.0	N/A	https://www.maxquant.org
DESeq2	Love et al.[81]	http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html
Other		
BaseScope LS Probe BA-Hs-RNU6ATAC-1zz-st-C1	Advanced Cell Diagnostics	1039918
BaseScope LS Probe PPIB-1zz	Advanced Cell Diagnostics	710178
BaseScope LS Probe DapB-1zz	Advanced Cell Diagnostics	701028
Basescope 2.5 LS Assay	Advanced Cell Diagnostics	323600
pRSV2-p120-AmpR XL10 (p120)	Zappia and Oshlack[82]	N/A
pCMV7.1 SCN4A frag (hSCN4A)	Zappia and Oshlack[82]	N/A
pLUX hSCN8A-minigene AmpR STbl3 (hSCN8a)	Zappia and Oshlack[82]	N/A

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Mark Rubin (mark.rubin@unibe.ch), and by the co-corresponding author Rahul Kanadia (Rahul.kanadia@uconn.edu).

Material availability
This study did not generate new unique reagents.

Data and code availability
- The bulk RNAseq and scRNA-seq data generated during this study have been submitted on the European Genome-phenome Archive under the accession EGAS00001005546. The mass spectrometry proteomics data have been deposited to the ProteomXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD026949. Original unprocessed images (western blot gels, RT PCR, and histology images) have been deposited at https://doi.org/10.17632/28j9zzdh3k.1. All datasets are publicly available as of the date of publication.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell and organoid lines
PNT1A (male, Sigma Aldrich, CB_95012614), LNCaP (male, ATCC, RRID: CVCL_1379), C4-2 (male, ATCC, RRID: CRL-3314), 22Rv1 (male, ATCC, RRID: CRL-2505), PC3 (male, ATCC, RRID: CRL-3470), DLD-1 (2 (male, ATCC, RRID: CCL-221), L-rENZ and L-AR cells were maintained in RPMI medium (Gibco, A1049101), supplemented with 10% FBS (Gibco, 10270106), and 1% penicillin-streptomycin (Gibco, 11548876) on poly-L-lysine coated plates. RWPE cells (male, ATCC, RRID: CVCL_3791) were maintained in Keratinocyte Serum Free Medium (Gibco, 17005075) supplemented with bovine pituitary extract and human recombinant EGF (included), and 1% penicillin-streptomycin (Gibco, 11548876). HEK293T cells (female, ATCC, RRID: CVCL_0063), VCaP (male, ATCC, RRID: CRL-2876), MDA-MB-231 (female, ATCC, RRID: HTB-26), K-562 (female, ATCC, RRID: CCL-243), LN-18 (male, ATCC, RRID: CRL-2610) PC-3M-Pro4 and DU145 cells (male, ATCC, RRID: CVCL_0105) were maintained in DMEM (Gibco, 31966021), supplemented with 10% FBS, and 1% penicillin-streptomycin. NCI-H660 cells (male, ATCC, RRID: CRL-5813) were maintained in RPMI medium (Gibco, A1049101), supplemented with 5% FBS, 1% penicillin-streptomycin (Gibco, 11548876), 0.005 mg/ml Insulin (Sigma-Aldrich, I9278), 0.01 mg/ml Apo-Transferrin (Sigma-Aldrich, T1147), 30nM Sodium selenite (Sigma-Aldrich, S9133), 10 nM Hydrocortisone (Sigma-Aldrich, H6909) 10 nM beta-estradiol (Sigma-Aldrich, E2257) and L-glutamine (for final conc. of 4 mM) (Sigma-Aldrich, G7513). PC-3M-Pro4 cells were a kind gift from Dr. Kruithof-De Julio. LNCaP-AR cells were a kind gift from Dr. Sawyers and Dr. Mu (Memorial

Sloan Kettering Cancer Center).[83] L-ENZ cells were established through constant enzalutamide exposure. Briefly low passaged LNCaP cells were treated over night with 20uM enzalutamide in C/S media. The media was exchanged to normal RPMI (10% FBRS, 1% P/S) the next day and surviving LNCaP cells (∼10%) were maintained until they reached a confluency of ∼80%. This procedure was repeated twice. Subsequently the enzalutamide concentration was increased for three treatments to 40uM and for 25 treatments to 80uM. Cells are treated since them once a week with 80uM enzalutamide.

All cell lines were grown at 37 °C with 5% CO_2. All cell lines were authenticated by STR analysis and regularly (every 3 month) tested for mycoplasma.

HP85 normal organoid line was a kind gift from Dr. Wouter Karthaus.[84] MSKCC-PCa8,10,14 and 16 CRPC-Adeno patient derived organoids were a kind gift from Dr. Chen[30] (Memorial Sloan Kettering Cancer Center). All organoids including PM154 were maintained in three-dimension according to the previously described protocol.[29,30] Briefly Advanced DMEM (Thermo Fisher Scientific, 31966047) with GlutaMAX 1x (Thermo Fisher Scientific, 35050061), HEPES 1mM (Thermo Fisher Scientific, 15630056), AA 1x (Life Technologies, 15240-062), 1% penicillin-streptomycin, B27 (Thermo Fisher Scientific,17504001), N-Acetylcysteine 1.25 mM (Sigma-Aldrich, A9165), Recombinant Murine EGF 50 ng/ml (PeproTech, 315-09), Human Recombinant FGF-10 20 ng/ml (Peprotech, 100-26), Recombinant Human FGF-basic 1 ng/ml (Peprotech, 100-18B), A-83-01 500 nM (Tocris, 29-391-0), SB202190 10 μM (Sigma-Aldrich, S7076), Nicotinaminde 10 mM (Sigma-Aldrich, N0636), (DiHydro) Testosterone 1 nM (Fluka, 10300), PGE2 1 μM (Tocris, 2296), Noggin conditioned media (5%) (PeproTech, 120-10C) and R-spondin conditioned media (5%) (PeproTech, 315-32). The final resuspended pellet was mixed with growth factor-reduced Matrigel (VWR, BDAA356239) in a 1:2 volume ratio. Droplets of 40 μl cell suspension/Matrigel mixture were pipetted onto each well of a six-well cell suspension culture plate (Huberlab, 7.657185) To solidify the droplets the plate was placed into a cell culture incubator at 37 °C and 5% CO_2 for 30 min. Subsequently 3 ml of human organoid culture media was added to each well. 50 % of the media was exchanged every 3−4 day during organoid growth. organoids were passaged as soon as they reached a size from 200 to 500 um. To this end, organoid droplets were mixed with TrypLE Express (Gibco) and placed in a water bath at 37 °C for a maximum of 5 min. The resulting cell clusters and single cells were washed and re-cultured, according to the protocol listed above.

In situ *validation collection*
Tissue micro-arrays were kindly provided by the Translational Research Unit (TRU) Platform, Bern (www.ngtma.com). For PCa we used TMAs from the Bern PCBM cohort[85] (28 patients) and a tissue microarray of 210 primary prostate tissues, part of the European Multicenter High Risk Prostate Cancer Clinical and Translational research group (EMPaCT).[86–88]

METHOD DETAILS

Mass spectrometry analysis
LNCaP, C4-2, 22Rv1 and PM154 cells (400 000) were seeded in a 6 well and treated for 96 hours with siScrambled or siU6atac RNA (16 pmol). 96 hours post transfection cells were harvested and 50% of the cell pellet was used for U6atac knockdown confirmation by qRT-PCR. The remaining pellet was washed twice with PBS and subjected to mass spectrometry (MS) analysis:

Cells were lysed in 8M urea/100mM Tris pH8 / protease inhibitors with sonication for 1 minute on ice with 10 seconds intervals. The supernatant was reduced, alkylated and precipitated overnight. The pellet was re-suspended in 8M urea/50mM Tris pH8 and protein concentration was determinate with Qubit Protein Assay (Invitrogen).10μg protein was digested with LysC 2hours at 37C followed by Trypsin at room temperature overnight. 800ng of digests were loaded in random order onto a pre-column (C18 PepMap 100, 5μm, 100A, 300μm i.d. x 5mm length) at a flow rate of 50μL/min with solvent C (0.05% TFA in water/acetonitrile 98:2).

After loading, peptides were eluted in back flush mode onto a home packed analytical Nano-column (Reprosil Pur C18-AQ, 1.9μm, 120A, 0.075 mm i.d. x 500mm length) using an acetonitrile gradient of 5% to 40% solvent B (0.1% Formic Acid in water/acetonitrile 4,9:95) in 180min at a flow rate of 250nL/min. The column effluent was directly coupled to a Fusion LUMOS mass spectrometer (Thermo Fischer, Bremen; Germany) via a nano-spray ESI source.

Data acquisition was made in data dependent mode with precursor ion scans recorded in the orbitrap with resolution of 120'000 (at m/z=250) parallel to top speed fragment spectra of the most intense precursor ions in the Linear trap for a cycle time of 3 seconds maximum.

Generation of U6atac and REST4 overexpressing cell lines
LV290591 – RNU6ATAC Lentiviral Vector (Human) (CMV) (pLenti-GIII-CMV-GFP-2A-Puro) as well as the corresponding empty vector control were purchased from ABM. pLV[Exp]-EGFP:T2A:Puro-CMV>[59]-REST4 Lentiviral vector and corresponding empty vector were purchased from vector builder. DNA was amplified via chemical transformation of One Shot Mach1 T1 Phage-Resistant Chemically Competent E. coli cells (Invitrogen, C862003). Lentivirus was produced in HEK293T cells by transfection with the constructs, and subsequent virus containing media was used to transduce C4-2 cells. Three days post transduction the cells were subjected to puromycin selection (1 μg/mL). After the selected cells reached a confluence of 80%, they were FACS sorted for GFP positivity. This was repeated 3 times (Figure S5C).

Drug treatments
For DHT stimulation experiments cells were starved of hormone for 48 hours in phenol red-free RPMI media (Gibco, 11-835-030) with 10% charcoal stripped FBS (Gibco, A3382101), then treated with 10 nM dihydrotestosterone (Fluka, 10300) for 24 hours.

For long-term ADT treatment cells were exposed weekly to 20 μM enzalutamide (Selleck Chemicals, S1250) or 10 uM Abiraterone.

For growth experiments cells were treated with siRNA and two hours later with 20uM enzalutamide. Enzalutamide was refreshed 3 days later.

For splicing reporter assays cells were exposed to Anisomycin 1ug/ml for 4h (Sigma Aldrich, A9789).

Cell transfection and siRNA-mediated knockdown

Cells

ON-TARGET plus siRNA SMARTpool siRNAs against *AR* (L-003400-00-00059), *EZH2* (L-004218-00-0005), *PDCD7* (L-012096-00-0005), *RNPC3* (L-063293-01-0005) and the non-targeting (siScrambled) siRNA were purchased from Dharmacon. siRNAs against RNU6atac (4390771, nn264646) and RNU12 (4390771, n265245)) and the Silencer Select Negative Control (4390843) were purchased from Thermo Fisher Scientific and siRNA against mouse U6atac was purchased from Ambion (4390827). Transfection was performed for the respective timepoints on attached cells using the Lipofectamine™ RNAiMAX Transfection Reagent (Thermo Fisher Scientific, 13778150) to the proportions of 16pmol of 20 μM siRNA per well.

Organoids

Before transfection organoids were cultured for 2-3 weeks in human organoid growth medium. Media was removed and organoids were first mechanically dissociated. To obtain single cells organoids were trypsinized in 1ml TriplE (Thermo Fisher Scientific, 12605036) for 15-18 minutes at 37C. The reaction was stopped with 1ml growth media and cells were spun for 5 minutes at 300g. Subsequently the cells were strained and counted. Per condition one million cells were plated in a 6 well. Lipofectamine™ RNAiMAX complexes were prepared according to the standard Lipofectamine™ RNAiMAX protocol. In short, 5ul of RNAiMAX reagent and 40 nM of siRNA plus 10% FBS were each diluted in 125 ul Opti-MEMH medium. Both mixes were pooled and incubated for 10 minutes before the siRNA-reagent complex was added to the cells. Cell/siRNA mix was centrifuged at 600 g at 32C for 60 min, and then incubated over night at 37C. The next day cells were resuspended and collected by centrifugation (300g, 5min, RT). The pellet was resuspended in 280 ul Matrigel and the mix was separated into 7 drops that were added into a 6 well. Organoids were grown in human organoid media for 96h (CTG assay) or seven days (cell counting assay).

RNA extraction from cells and qRT-PCR

Cells were harvested for RNA isolation using the ReliaPrep™ miRNA Cell and Tissue Miniprep System (Promega, Z6212). Synthesis of complementary DNAs (cDNAs) using FIREScript RT cDNA Synthesis Kit (Solis BioDyne, 06-15-00200) and real-time reverse transcription PCR (RT-PCR) assays using HOT FIREPol EvaGreen qPCR Mix Plus (Solis BioDyne, 08-24-00020) were performed using and applying the manufacturer protocols. Quantitative real-time PCR was performed on the ViiA 7 system (Applied Biosystems). All quantitative real-time PCR assays were carried out using three technical replicates. Relative quantification of quantitative real-time PCR data used GAPDH, ACTB as housekeeping genes. Primer sequences are listed in Table S14.

RT-PCR

Gel-based RT-PCR against MORC4, XPO4, SPSC and SNRSF10 was performed. The rational here is that those four MIGs showed the strongest increase in AS events upon siU6atac treatment. RNA was isolated and 1ug RNA was reverse transcribed into cDNA as described above. RT-PCR was performed using Platinum™ SuperFi II Green PCR Master Mix (Thermo Scientific, 12369010) and 60ng cDNA template. Primers were designed to detect predicted AS events (Table S14). RT-PCR product was loaded on a 1% agarose gel and visualized. MSI values were determined by quantification of band intensity by ImageJ.[89] Formula: Total AS events divided by total events multiplied by 100.

Single-cell sequencing

Cell counting and viability assessments were conducted using a ViCell XR Cell counter and viability analyzer (Beckman Coulter, BA30273). Thereafter, GEM generation & barcoding, reverse transcription, cDNA amplification and 3' gene expression library generation steps were all performed according to the Chromium Next GEM Single Cell 3' Reagent Kits v3.1User Guide (10x Genomics CG000204 Rev D) with all stipulated 10x Genomics reagents. Generally, 11.8.-27.5 μL of each cell suspension (600-1'400 cells/μL) and 15.7-31.4 μL of nuclease-free water were used for a targeted cell recovery of 10'000 cells. GEM generation was followed by a GEM-reverse transcription incubation, a clean-up step and 10-12 cycles of cDNA amplification. The resulting cDNA was evaluated for quantity and quality using a Thermo Fisher Scientific Qubit 4.0 fluorometer with the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Q32854) and an Advanced Analytical Fragment Analyzer System using a Fragment Analyzer NGS Fragment Kit (Agilent, DNF-473), respectively. Thereafter, 3' gene expression libraries were constructed using a sample index PCR step of 11-12 cycles. The generated cDNA libraries were tested for quantity and quality using fluorometry and capillary electrophoresis as described above. The cDNA libraries were pooled and sequenced with a loading concentration of 300 pM (150 pM in runs using XP workflow), paired end and single indexed, on an illumina NovaSeq 6000 sequencer using a NovaSeq 6000 S2 Reagent Kit v1.5 (100 cycles; illumina 20028316) and two NovaSeq 6000 S4 Reagent Kits v1.5 (200 cycles; illumina 20028313). The read set-up was as follows: read 1: 28 cycles, i7 index: 8 cycles, i5: 0 cycles and read 2: 91 cycles. The quality of the sequencing runs was assessed using illumina Sequencing Analysis Viewer (illumina version 2.4.7) and all base call files were demultiplexed and converted into FASTQ files using illumina bcl2fastq conversion software v2.20. All steps were performed at the Next Generation Sequencing Platform, University of Bern.

Bulk RNA sequencing

Total RNA was extracted from LNCaP, C4-2, 22Rv1 and PM154 cells treated for 96h with siU6atac or siScrambled. The recommended DNase treatment was included. The quantity and quality of the extracted RNA was assessed using a Thermo Fisher Scientific Qubit 4.0 fluorometer with the Qubit RNA BR Assay Kit (Thermo Fisher Scientific, Q10211) and an Advanced Analytical Fragment Analyzer System using a Fragment Analyzer RNA Kit (Agilent, DNF-471), respectively. Thereafter cDNA libraries were generated using an illumina Stranded Total RNA Prep, Ligation with Ribo-Zero Plus (illumina, 20040529) in combination with IDT for Illumina RNA UD Indexes Sets A and B (Illumina, 20040553 and 20040554, respectively). The illumina protocol was followed exactly with the recommended input of 100 ng total RNA. The quantity and quality of the generated NGS libraries were evaluated using a Thermo Fisher Scientific Qubit 4.0 fluorometer with the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Q32854) and an Advanced Analytical Fragment Analyzer System using a Fragment Analyzer NGS Fragment Kit (Agilent, DNF-473), respectively. As a further quality control step, prior to NovaSeq 6000 sequencing, the pooled cDNA library pool underwent paired end sequencing using iSeq 100 i1 reagent v2, 300 cycles (illumina, 20040760) on an iSeq 100 sequencer. The library pool was re-pooled to ensure an equal number of reads/library and then paired end sequenced using a NovaSeq 6000 S4 reagent kits v1.5, 300 cycles (illumina, 20028312) on an Illumina NovaSeq 6000 instrument. The quality of the sequencing runs was assessed using illumina Sequencing Analysis Viewer (illumina version 2.4.7) and all base call files were demultiplexed and converted into FASTQ files using illumina bcl2fastq conversion software v2.20. The average number of reads/ libraries was 82 million. The RNA quality-control assessments, generation of libraries and sequencing runs were performed at the Next Generation Sequencing Platform, University of Bern, Switzerland.

U6atac *in situ* hybridization

mRNA ISH was performed by automated staining using Bond RX (Leica Biosystems) and Basescope® technology (Advanced Cell Diagnostics, Hayward, CA, USA). All slides were dewaxed in Bond dewax solution (product code AR9222, Leica Biosystems) and heat-induced epitope retrieval at pH 9 in Tris buffer based (code AR9640, Leica Biosystems) for 15 min at 95° and Protease treatment for 5 min. The following probes from RNAscope 2.5 LS (Advanced Cell Diagnostics) were used: BaseScope™ LS Probe - BA-Hs-RNU6ATAC-1zz-st-C1ref 1039918, PPIB-1zz ref 710178 and DapB-1zz ref 701028, were used as positive and negative control respectively. Probe efficiency was tested using U6atac overexpressing C4-2 cells (5 million) of which 50% were treated with siU6atac RNA.

All probes were incubated at 37° for 120 min. Basescope™ 2.5 LS Assay (Ref 323600, Advanced Cell Diagnostics) was used as pre-amplification system. Subsequent the reaction was visualized using Fast red as red chromogen (Bond polymer Refine Red detection, Leica Biosystems, Ref DS9390) for 20 min. Finally, the samples were counterstained with Haematoxylin, air dried and mounted with Aquatex (Merck). Slides were scanned and photographed using Pannoramic 250 (3DHistech). U6atac intensity was scored manually by a pathologist (Mark Rubin) blinded to the clinical data, using the digital online TMA scoring tool Scorenado[90] (University of Bern, Switzerland) especially developed for TMA scoring on de-arrayed spots.

For the analysis of TMA data, samples annotated as 'center' were used. U6ATAC score was calculated by multiplying the percentage of positive cells by the intensity. The sample with the highest score was used where more than one value was recorded for a block. Comparisons between groups were carried out using Wilcox test.

325 primary, 25 primary with metastatic potential and 32 metastatic samples, from 24 patients were used for the comparison of U6ATAC expression in PCa and PCBM.

Flow cytometry

C4-2 and PM154 cells were seeded in a 6 well (500 000/well) and transfected with siU6atac or siScrambled RNA for 72 and 96 hours as previously described. Flow Cytometry cell cycle analysis was performed using the Click-iT™ EdU Alexa Fluor™ 488 Flow Cytometry Assay Kit (Thermo Fisher Scientific, C10420). Briefly EdU (10uM) was added into the media and cells were incubated for one hour at 37C. cells were washed with 1% BSA in PBS and fixed in 100ul Click-iT fixative for 15 minutes. After three additional washing step cells were permeabilized for 15 minutes in 100ul 1xClick-iT saponin based reagent. Click-iT reaction cocktail was prepared according to manufacturer's instructions and 500ul reaction mix/ condition were incubated for 30 min with the cells at room temperature. Cells were washed and resuspended in 500 ul saponin-based permeabilization buffer. Hoechst (1ug/ml) was added 20 minutes prior analysis to the reaction mix. Cells were analyzed using the FACSDiva Software on a BD LSR II Flow Cytometer (BD Biosciences) in the FACSlab Core facility of the University of Bern. Data was further quantified with FlowJo 10.7.1. Values were calculated as fold-change as compared to siScrambled treated controls.

Immunoblotting

Cells were lysed in GST-Fish buffer (10 % (v/v) Glycerol, 50 mM Tris-HCl pH7,4, 100 mM NaCl, 1 % (v/v) Nonidet P-40, 2 mM MgCl₂, 1 mM PMSF) with freshly added protease and phosphatase inhibitors. Total protein concentration was measured using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). 50 ug protein samples were resolved 4-15% Mini-Protean TGX gels (BioRad, 456-1084) in SDS-PAGE and transferred to nitrocellulose membranes using the iBlot2 system (Thermo Fisher, IB23001). Blots were blocked for 1 hour at room temperature in 5% milk/TBST or BSA/TBST and incubated overnight at 4 °C with primary antibodies (Table S14) which were dissolved in 5% BSA/TBST buffer. After 3 washes, the membrane was incubated with secondary antibody conjugated to horseradish peroxidase for 1 h at room temperature. After 3 washes, signal was visualized by chemiluminescence

using the Luminata Forte substrate (Thermo Fisher Scientific, WBLUF0100) for strong antibodies and WesternBright Sirius-HRP Substrate (Witec AG, K-12043-D10) for weak antibodies. Images were acquired with the FUSION FX7 EDGE Imaging System (Witec AG).

Luciferase reporter assay

CMV-luc2CP/ARE (major intron splicing reporter), CMV-luc2CP (empty vector backbone control) and luc1CFH4 (minor splicing intron reporter) were a kind gift from Dr. Gideon Dreyfuss[28] (University of Pennsylvania). DNA was amplified via chemical transformation of One Shot Mach1 T1 Phage-Resistant Chemically Competent E. coli cells (Invitrogen, C862003) and Sanger sequenced.

To determine the minor/major intron splicing rate cells were seeded in white 96 well plate (Huberlab, 7.655 098) (8000/well) and treated according to assay conditions. 24 hours prior analysis cells of each condition were co-transfected with each reporter plasmid and the empty vector backbone plasmid. In short 1.5ul of P3000 reagent plus 0.5 ug of DNA and 1.5 ul Lipofectamine P300 were each diluted in 25ul Opti-MEMH medium. Both mixes were pooled and incubated for 20 minutes before the solution was added to the cells. Luciferase expression was measured with the Dual-Glo® Luciferase Assay System (Promega, E2940): media was removed, 100ul of PLB were added and cells were frozen for two hours at -20C. After a one hour shaking step 100 ul of LAR substrate was added and Firefly luciferase expression was measured with a Tecan Infinite M200PRO reader. Values were calculated as x-fold of CMV-luc2CP expression and subsequently as the x-fold of the respective reference control.

Minigene reporter assay

pRSV2-p120-AmpR XL10 (p120), pCMV7.1 SCN4A frag (hSCN4A) and pLUX hSCN8A-minigene AmpR STbl3 (hSCN8a) were a kind gift from Dr. Mark-David Ruepp[71] (King's College London). DNA was amplified via chemical transformation of One Shot Mach1 T1 Phage-Resistant Chemically Competent E. coli cells (Invitrogen, C862003).

To determine the minigene splice index cells were seeded in a 6-well (350 000/well) and treated according to assay conditions. SiRNA was added 96 hours prior measurement. 72 hours prior measurement cells were transfected with the respective minigene. Briefly 10 ul of P3000 reagent plus 1.5 ug of DNA and 7.5 ul Lipofectamine P3000 were each diluted in 125ul Opti-MEMH medium. Both mixes were pooled and incubated for 20 minutes before the solution was added to the cells. 48 hours before the measurement media was exchanged for media with 10% charcoal stripped FBS (Gibco, A3382101) and 24 hours prior measurement 100 nM DHT was added to the respective condition. qRT-PCR was performed to verify the knockdown and to determine the splice index of each minigene. The minigene splice index was calculated by forming the ratio of normalized mRNA levels of cells transfected with the minigene versus mRNA levels of WT cells to consider the transfection efficiency. Subsequently the values corresponding to the spliced minigene were divided by the values corresponding to the unspliced minigene.

Cell-growth experiments

Viability

Cells were seeded in a 6 well (400 000) and treated according to assay conditions over night. Cells were then seeded in Poly-L-Lysine coated 96-well plates (8000 cells/well, n=3 per condition) and PM154, MSK10 and MSK16 organoids were seeded in a collagen-coated 96-well plates to allow for 2D growth (5000 cells/well, n=3 per condition). Remaining cells were used for U6atac knockdown control via qRT-PCR. Cell viability was determined after 24, 48, 72, and 96 h with a Tecan Infinite M200PRO reader using the CellTiter-Glo® Luminescent Cell Viability Assay according to manufacturer's directions (Promega, G9243). Viability values were calculated as x-fold of cells transfected with siRNA for 0 h. Cell confluence (n=4 per condition) was determined using the Incucyte S3 instrument and the IncuCyte S3 2018B software (Essen Bioscience, Germany). Values were calculated as x-fold of timepoint 0 and then as fold-change in confluency as compared to siScrambled treated controls.

Organoids

Organoids were transfected with siRNA as described previously. The following day 160000 cells were resuspended in 320 ul Matrigel and drops of 40ul (one drop/well, four timepoints, n=2) were plated in a suspension 48-well plate (Huberlab, 7.677 102). Remaining cells were plated in a 6-well for q RT-PCR U6atac knockdown control. The 24-well plate was incubated for three minutes in 37C and for 20 minutes upside down in 37C. Subsequently 500ul of organoid media were added and viability was measured using the CellTiter-Glo 3D Cell Viability Assay (Promega, G9683).

Doubling time experiment

Incucyte experiments were performed with HP85, MS2514, DLD1, LN18, Hek293, L-AR, K562, MDA231, C4-2, 22Rv1, LNCaP, Hs27 and PM154 cells. Briefly, confluence of cells treated with siU6atac or siScrambled was measured every six hours for a total of 120h. At the end of each Incucyte experiment viability of every sample was tested by cell titer glow (CTG) assays. Incucyte, and CTG results (y-axis) were calculated as x-fold of the scrambled control and plotted against the doubling time (DT) of each cell line (x-axis). Cell lines were ranked according to their doubling time.

Co-culture experiments

Co-culture experiments were performed with C4-2-GFP cancer cells and HS27-mCherry fibroblasts Briefly, C4-2 cells were fluorescently tagged with 0.5nM green fluorescent AIEDot nanoparticles (Merck, SCT012), while HS27 cells were tagged with 0.5nM red fluorescent AIEDot nanoparticles (Merck, SCT011). Subsequently, both cell lines were co-cultured in a one (HS27) to three (C4-2) ratio (total 400 000 cells) and treated with Scrambled or U6atac siRNA in a 6well plate. Cell growth was measured every six hours

for a total of 5 days using the Incucyte S3 instrument and the IncuCyte S3 2018B software (Essen Bioscience, Germany). Values were calculated as x-fold of timepoint 0 and then as fold-change in confluency as compared to siScrambled treated controls. At the end of the experiment, the cell mix was FACS sorted to separate C4-2, and HS27 cells and U6atac knockdown was confirmed via FACS sorting and subsequent qRT-PCR.

QUANTIFICATION AND STATISTICAL ANALYSIS

Protein-protein interaction network analysis

Assuming that cancer genes perturb a large molecular network through their interactions with other genes, distance to MIGs in a protein-protein interaction network of 160,881 interactions between 15,366 human proteins as of the HINT database was determined.[72] In particular, a MIG that directly interacts with a cancer-causing gene is a distance $d=1$ away, while a protein that is separated by a protein in between is a distance $d=2$ away from a cancer-genes in question. To find these interaction distances, a list of 403 cancer-causing genes from 186 different cancer types as of the Cancer Genome Interpreter database was considered (Table S1) (https://www.cancergenomeinterpreter.org)[20] in the underlying protein-protein interaction network of 160,881 interactions between 15,366 human proteins as of the HINT database.[72] A profile of the numbers of MIGs with a given distance d away from each cancer gene, N_d^{MIG} ($d = 1, …, 5$) was obtained. To assess if the presence of MIGs in the vicinity of cancer genes is significant, sets of 542 MIGs found in the underlying interaction networks were randomly sampled. In particular, the corresponding distances of cancer genes to these randomly sampled MIGs was measured and a profile of numbers of random MIGs with a given distance d away from cancer genes $N_d^{r,MIG}$ ($d = 1, …, 5$) was obtained. The enrichment of MIGs is defined a distance d away from cancer genes as $E_d^{MIG} = lg_2\left(\frac{N_d^{MIG}}{N_d^{r,MIG}}\right)$, with average E over 100,000 random samples of MIGs.

To find a tight-knit web of MIGs and cancer genes, such direct interactions of cancer genes and MIGs were distracted in the underlying human protein-protein interaction network and determined the size of the largest connected subnetwork S. Randomly sampling MIGs N = 100,000 times. The sizes of the largest subnetworks were calculated using these random sets, S_r, and the significance of S by $P = \frac{\#(S_r > S)}{N}$ was determined.

Principal component analysis

MIG expression data from 3 sources: prostate samples from GTEx (healthy prostate tissue), prostate cancer samples from TCGA (primary prostate cancer samples), and prostate cancer samples from SU2C (advanced prostate cancer) was merged. Gene expression values were normalized following the protocol adopted by GTEx (and detailed in our pan-cancer analysis). PCA analysis on this normalized gene expression matrix was carried out, thereby enabling visualization of the resultant data as the projections onto the space spanned by the first 2 PCs. Notably, this visualization appears to capture the progression from these 3 broad phases of prostate cancer progression, from healthy tissue in GTEx (at lower ends of the first PC) to advanced stages in SU2C (at higher ends of the first PC).

Quantitative comparisons between MIG- and non-MIG-based gene expression clustering

The Silhouette coefficient[73] was used in order to characterize the relative performance of gene expression clustering for gene sets containing different relative abundances of MIGs and non-MIGs. The Silhouette coefficient provides an objective metric for measuring what is visually discerned to be structure (or any lack thereof) in a given heatmap (Figure S1; Table S2). This coefficient constitutes an unsupervised approach to provide a score ranging -1 and 1, with scores closer to 1 indicative of well-defined and dense clustering (i.e., more meaningful structure in a given heatmap). This coefficient quantifies how similar a given data point (i.e., sample) is to its own cluster relative to different clusters. In our study, a data point consists of a N-length vector, where N is the number of distinct samples in an expression matrix, and the number of distinct clusters is pre-defined to be n_cluster = 23 in our pan-cancer analysis, since this analysis was carried out on a dataset of 23 distinct cancer types. The Silhouette coefficient has also been adopted for similar purposes in previous studies[22–25]

Prior to calculating the Silhouette coefficient, an agglomerative clustering on a normalized gene expression matrix was performed. Gene expression values from 23 cancer types (Biliary-AdenoCA, Bladder-TCC, Bone-Leiomyo, Breast-AdenoCA, Cervix-SCC, CNS-GBM, CNS-Oligo, ColoRect-AdenoCA, Head-SCC, Kidney-ChRCC, Kidney-RCC, Liver-HCC, Lung-AdenoCA, Lung-SCC, Lymph-BNHL, Lympg-CLL, Ovary-AdenoCA, Panc-AdenoCA, Prost-AdenoCA, Skin-Melanoma, Stomach-AdenoCA, Thy-AdenoCA, Uterus-AdenoCA) (Table S2) totaling N=1224 samples were taken from PCAWG, and the gene expression normalization was performed using the same approach as that adopted by GTEx (PMID: 29022597). Briefly, the entire gene expression matrix was normalized using quantile normalization. Then, inverse quantile normalization was applied to this quantile-normalized matrix in order to map to a standard normal (this also enabled us to remove outliers). For the PCa progression analysis, gene expression values from GTEx, TCGA, and SU2C were merged into one expression matrix (Figure S1). Briefly the pre-normalized expression matrix was normalized using quantile normalization, followed by inverse quantile normalization to map to a standard normal distribution; this step also removed outlier genes. This normalization scheme was the same as that adopted previously by GTEx.[70] Using this normalized gene expression matrix, clustering was performed by using hierarchical clustering by employing the Euclidean metric and Ward linkage.

The entire analysis was run using different expression matrices with varying fractions of MIG genes, and the results are plotted in Figures 1F and 1G. Each box in this plot (for instance, the right-most box, which represents 100% non-MIG sets) corresponds to a 1000-length simulation in which non-MIGs are randomly sampled from among all non-MIGs in the genome. Thus, for the case of the right-most box, 1000 random sets are sampled, each of which has a composition of 100% non-MIGs. The P-value is based on a two-sided t-test in which each sample value is taken to be the difference between the Silhouette coefficient of one of the 1000 100% non-MIG samples and the corresponding Silhouette coefficient when only MIGs are used for clustering (i.e., the Silhouette coefficient corresponding to 0% non-MIGs).

MSI analysis

Primary tumor RNA-seq patient samples from The Cancer Genome Atlas (TCGA) were randomly queried using the GenomicDataCommons package in R (http://github.com/Bioconductor/GenomicDataCommons). The minor intron retention pipeline developed by Olthof et al. (https://github.com/amolthof/minor-intron-retention)[19] was run on the queried TCGA samples from the following cohorts: breast invasive carcinoma (BRCA; N=20), cholangiocarcinoma (CHOL; N=20), colon adenocarcinoma (COAD; N=20), lung adenocarcinoma (LUAD; N=20), ovarian serous cystadenocarcinoma (OV; N=20), pancreatic adenocarcinoma (PAAD; N=20), prostate adenocarcinoma (PRAD; N=16), and thyroid carcinoma (THCA; N=20). For the prostate cancer analyses, the following additional samples from the Stand Up to Cancer dataset were analyzed: androgen receptor (AR; N=20) and neuroendocrine (NE; N=22). NE samples were samples which had an NEPC score of greater than 0.4 while AR samples had an NEPC score of less than or equal to 0.4. Prostate cancer samples from the Genome-Tissue Expression Portal were analyzed as well (GTEX; N=20). A Kruskal-Wallis with post-hoc Dunn's Test was performed between the TCGA cohorts and the prostate cohorts. A heatmap was generated with the gplots package in R with the default clustering method for the pan-cancer TCGA cohorts and the prostate cohorts across the MIGs (https://CRAN.R-project.org/package=gplots). A GO Enrichment Analysis was performed on the genes that clustered for each cancer cohort, grouping genes with a mean MSI value in the ranges of 0, 0 to 0.04, 0.04 to 0.75, and 0.75 to 1.

RNU11 quantification according to Gleason score

Gene-expression data of primary prostate cancer specimen was retrieved from The Cancer Genome Atlas (TCGA) in form of raw-counts. Sequencing reads were aligned to the human reference genome (hg38) using *STAR*.[91] Gene-expression was quantified at gene-level using *Gencode* (https://www.gencodegenes.org) annotations (v29).[92] Subsequent analysis and library-size normalization were performed using *edgeR* pipeline.[93] RNU11 mapping reads were identified in 23 out of 497 samples (5%) which reflects the difficulty of capturing this gene product using canonical PolyA+ sequencing techniques. Nonetheless, a clear association between Gleason score and RNU11 mRNA expression in these 23 samples was identified (Table S5). Significance was assessed using non-parametric Wilcoxon Test.

Single-cell RNAseq

Single-cell preprocessing and quality control

Cell ranger analysis pipeline v6.0.1 was used to align reads to the human genome reference sequence (GRCh38) and generate a gene-cell matrix from these data. The gene expression matrix was analyzed using Seurat 4.0.3 (https://github.com/satijalab/seurat).[94] We removed low quality cells and multiplets by excluding genes detected in less than 5 cells and by discarding cells with more than 10000 fewer than 1000 detected genes. Cells containing mitochondrial gene counts greater than 25% were also removed.

UMI counts were normalized with the NormalizeData Seurat function using the LogNormalize normalization method with default parameters (10000 scale.factor).

Cell-cycle phase classification and cell scores

Prediction of cell cycle phase for each cell was performed using Seurat CellCycleScoring function. A score was computed and a cell phase (G2/M, S and G1) was assigned to the cell as described previously.[49] Fisher's exact test was performed to check whether the sIU6 cells have significantly a different number of cells than the Scr in G1 or S phase using the R fisher.test function.

We used Seurat AddModuleScore function to evaluate the degree to which individual cells express a certain pre-defined gene set. We defined scores to estimate the activities of prostate AR pathway, and EMT state, as described previously.[50] The AR pathway gene set included *AR, KLK3, KLK2, FKBP5, TMPRSS2, FOXA1, GATA2, SLC45A3* and EMT state *CDH2, CDH11, FN1, VIM, TWIST1, SNAI1, ZEB1, ZEB2* and *DCN*. Violin plots were drawn using Seurat and *p-value*s were calculated using Wilcoxon test.[95]

Multiple datasets integration and batch correcting

For merging multiple datasets and minimizing the batch effect between them, we integrated our 6 samples (3 replicates SCR and 3 replicates sIU6) for each cell line following the procedure of Seurat v4.0.3.[96]

Briefly, we selected the most variable genes for each dataset using the FindVariableFeatures function (selection.method ="vst") and ranked them according to the number of datasets in which they were independently identified as highly variable. The 2000 most variable genes were thus integrated by merging pairs of datasets according to a given distance.

Integration anchors, representing two cells that are predicted to originate from a common biological state in both datasets using a Canonical Correlation Analysis (CCA), were done using the FindIntegrationAnchors function. The expression of the target dataset was corrected using the difference in expression between the two expression vectors for each pair of anchor cells. This step was

performed using the IntegrateData function. This process resulted in an expression matrix containing the batch-effect-corrected expression for the 2000 selected genes for all cells from the 6 samples for each cell line.

Dimension reduction and clustering

A PCA was performed on the scaled data using RunPCA Seurat function (npcs = 30). Uniform Manifold Approximation and Projection (UMAP), a nonlinear dimension reduction method, was run using RunUMAP Seurat package function in order to embed cells in a 2-dimensional space. A K-nearest neighbor graph (KNN) based on the Euclidean distance in PCA space was constructed to cluster the cells with the Louvain algorithm (resolution = 0.2) using the FindNeighbors and FindClusters Seurat functions. We selected the optimal clustering resolution using the clustree R package (v0.4.3).[82] Barplots were performed using dittoSeq R package.[97]

Differential gene-expression analysis

Differentially expressed genes (DEGs) were identified between the different clusters using the FindAllMarkers function from the Seurat package (one-tailed Wilcoxon rank-sum test, *p* values adjusted for multiple testing using the Bonferroni correction). To compute DEGs, all genes were tested provided they were expressed in at least 25% of cells in either of the two compared populations, and the expression difference on a natural log scale was at least 0.25. The heatmap was produced using the DoHeatmap Seurat function by selecting the top five genes for each cluster.

RNAseq

Gene expression

Paired-end, total RNA reads for each replicate (N=4) were mapped to the mm10 genome via Hisat2.[98] Reads mapping to multiple locations were removed. Gene expression values were calculated using IsoEM2.[74] Differential gene expression was calculated by IsoDE2,[74] which uses 200 rounds of iterative bootstrapping to produce a 95% confidence interval for the expression of each gene, then statistically compares these values between experimental conditions. A threshold of $\log_2 FC \geq 1$, $P \leq 0.01$ for upregulation, and $\log_2 FC \leq -1$, $P \leq 0.01$ for downregulation was employed.

Minor intron retention

We report here minor intron retention as a mis-splicing index through the methodology described in Olthof et al..[19] Briefly, uniquely mapped reads from the region of interest around minor introns (from two exons upstream to two exons downstream) were extracted. The mis-splicing index was then calculated by summing reads that map to the 5' splice site and 3' splice site of a minor intron, divided by the sum of reads that map to the 5' splice site, 3' splice site, and 2x canonically spliced reads. We only considered introns that pass our filtering criteria, which requires >4 exon-intron boundary reads, ≥ 1 read mapping to both the 5' splice site and 3' splice site, and >95% intron coverage, in all replicates of a condition as retained. Statistically significant global minor intron retention was determined using a Kruskal-Wallis test with post-hoc Dunn's test ($P \leq 0.05$). Determination of individual MIGs with significantly elevated minor intron retention was calculated using a two-tailed student's T-test ($P \leq 0.05$)

To assess global intron retention, we used IRFinder[75] for each cell line using default parameters. Intron retention was considered significantly elevated if $\log_2 FC > 1$ and BH-adjusted p<0.05.

Alternative splicing

We employed the methodology reported by Olthof et al.[48] for our alternative splicing analysis. Briefly, we used BEDTools to classify differential 5' splice site and 3' splice site usage around the region of interest for all minor introns and binned them into one of 8 categories (Figure 3B). We then calculated a mis-splicing index by quantifying the number of AS reads divided by the sum of AS reads and canonically spliced reads per category per sample. We used a filtering criterion wherein we only considered introns to have AS if the average mis-splicing index for all replicates for each condition was > 10%. Additionally, we normalized the number of reads supporting an AS event by the total sequencing depth. As such, we only included AS events with >1 read per 3 million uniquely mapped reads for analysis. Determination of individual MIGs with significantly elevated AS was calculated using a two-tailed student's T-test ($P \leq 0.05$).

Alternative splicing analysis of major introns upstream (676) or downstream (657) of minor introns in MIGs was conducted as described previously in Olthof et al.[48] Similarly, the alternative splicing of 676 random major introns in non-MIGs was analyzed. AS events occurring only in the siScr or siU6atac condition were binned as downregulated or upregulated, respectively. Differential AS for events detected in both samples was assessed using student's two-tailed T-test, with p<0.05 used as cut-off for significance. In addition, we projected the AS mis-splicing index (MSI) for all AS events detected in either the siScr or siU6atac condition for each cell line as box-plots. We then tested whether the median AS-MSI was significantly different in the siU6atac condition using Mann Whitney U-Test with p<0.05 cut-off for significance

To link minor intron retention with proteomic changes, we predicted whether retention would trigger non-sense mediated decay (NMD) as we have done previously.[48] Specifically, we used a widely employed classification strategy where the introduction of a premature stop codon >50 nt upstream of the last exon-exon junction complex was predicted to trigger NMD.

DAVID analysis

Gene lists were submitted to DAVID for gene ontology (GO) enrichment.[99] We considered only GO Terms with Benjamini-Hochberg adjusted *P*-value ≤ 0.05 as significant.

STRING and Ingenuity Pathway Analysis

For LNCaP. C4-2, and 22Rv1, overlapping MIGs with significantly elevated minor intron retention that were also found to be associated with prostate cancer-causing genes were grouped with overlapping protein coding genes with significant downregulation.

This list was submitted to STRING[76] under the default parameters to obtain the gene interaction network. Subsequently, the same list was submitted to IPA[76] as a core analysis using default parameters. All reported biological networks and pathways from IPA were significant using a Benjamini-Hochberg adjusted P-value $\leq$ 0.05 cut-off. The same analysis was performed for PM154 alone.

Principal component analysis
Principal component analyses were performed using clustvis[100] with default parameters. Ellipses show 95% confidence interval.

GSEA analysis
To perform gene set enrichment analysis RNAseq data was pre-ranked using the metric: $\log_{10}$(Pvalue) / sign (logFC). GSEA was performed using the GSEA v.4.0.3 software.[101] Hallmark gene sets, obtained from the GSEA website (www.broadinstitute.org/gsea/) was used for enrichment of siU6atac related pathway genes. Dotplot was used to visualize the most significant enriched terms. Normalized enrichment score (NES) and False discovery rate (FDR) were applied to sort siU6atac pathway enrichment after gene set permutations were performed 1000 times for the analysis.

MassSpec
MS data was interpreted with MaxQuant (version 1.6.1.0) against a SwissProt human database (release 2019_07) using the default MaxQuant settings, allowed mass deviation for precursor ions of 10 ppm for the first search and maximum peptide mass of 5500 Da; match between runs with a matching time window of 0.7 min was activated, but prevented between different groups of replicates by the use of non-consecutive fractions. Furthermore, the four cell lines were treated as different parameter groups and normalized independently. Settings that differed from the default also included: strict trypsin cleavage rule allowing for 3 missed cleavages, fixed carbamidomethylation of cysteines, variable oxidation of methionines and acetylation of protein N-termini.

Protein intensities are reported as MaxQuant's Label Free Quantification (LFQ) values, as well as Top3 values (sum of the intensities of the three most intense peptides); for the latter, variance stabilization was used for the peptide normalization. Missing peptide intensities were imputed in the following manner, provided there was at least one identification in the group: two missing values in a group of replicates would be replaced by draws from a Gaussian distribution of width 0.3 x sample standard deviation centered at the sample distribution mean minus 1.8× the sample standard deviation, whereas a single missing value per group would be replaced following the Maximum Likelihood Estimation (MLE) method. Imputation at protein level for both LFQ and Top3 values was performed if there were at least two measured intensities in at least one group of replicates; missing values in this case were drawn from a Gaussian distribution of width 0.3 x sample standard deviation centered at the sample distribution mean minus 2.5x the sample standard deviation. Differential expression tests were performed using the moderated t-test empirical Bayes (R function EBayes from the limma package version 3.40.6) on imputed LFQ and Top3 protein intensities. The Benjamini-Hochberg method was further applied to correct for multiple testing. The criterion for statistically significant differential expression is that the maximum adjusted p-value for large fold changes is 0.05, and that this maximum decrease asymptotically to 0 as the $\log_2$ fold change of 1 is approached (with a curve parameter of one time the overall standard deviation). The protein imputation step was repeated 20x so as to be able to flag those proteins that are persistently significantly differentially expressed throughout the cycles.

Imputed iTop3 was used to calculate relative protein abundances. Differential expression was calculated using the Empirical Bayes test. Protein upregulation and downregulation was determined by setting a threshold of Benjamini-Hochberg adjusted P-value $\leq$ 0.05, $\log_2$FC $\geq$ 1 or $\leq$ -1, respectively.

REST/REST4 correlation analysis
Paired-end transcriptome-sequencing reads were retrieved from metastatic castration resistant prostate cancer dataset (mCRPC) (SU2C project)[55] by selecting samples obtained with the PolyA library method. Raw reads were quantified using Kallisto pseudoalignment[77] to the reference human genome GRCh38. Raw pseudocounts were normalized using DESeq2[78] and log transformed using the log function implemented in R (version 4.0.5). AR and NEPC scores were downloaded from cbioportal (https://www.cbioportal.org/study/summary?id=prad_su2c_2019). Scatterplots were drawn using ggplot2 and a Pearson correlation was calculated using the stat_cor function implemented in ggpubr R package (https://rpkgs.datanovia.com/ggpubr/). Boxplots were drawn using ggplot2 by selecting only the samples with AR or NEPC scores higher than 0.4[21] and a Wilcoxon test was performed using wilcox.test function implemented in ggsignif package.

PWM score calculation
REST intron classification was done using PWMs as described previously by us.[79] Briefly human genome and intron data was extracted from FASTA and GTF files obtained from Ensembl. Introns were binned as putative major, putative minor, or 'other' based on their terminal dinucleotide sequence, PPT and BPS. Initial PWMs were generated and all introns were scored and re-binned to generate refined PWMs for the major and minor 5'SS, minor BPS and major PPT.