







# An integrated functional and clinical genomics approach reveals genes driving aggressive metastatic prostate cancer

Rajdeep Das <sup>1,2</sup>, Martin Sjöström <sup>1,2</sup>, Raunak Shrestha<sup>1,2</sup>, Christopher Yogodzinski<sup>2,3</sup>, Emily A. Egusa<sup>1,2</sup>, Lisa N. Chesner<sup>1,2</sup>, William S. Chen<sup>1,2</sup>, Jonathan Chou <sup>2,4</sup>, Donna K. Dang<sup>1,2</sup>, Jason T. Swinderman<sup>2,3</sup>, Alex Ge<sup>2,3</sup>, Junjie T. Hua<sup>1,2</sup>, Shaheen Kabir<sup>2,3</sup>, David A. Quigley<sup>2,3,5</sup>, Eric J. Small <sup>2,4</sup>, Alan Ashworth<sup>2,4</sup>, Felix Y. Feng <sup>1,2,3,4,7,8</sup> ✉ & Luke A. Gilbert <sup>2,3,6,7,8</sup> ✉

Genomic sequencing of thousands of tumors has revealed many genes associated with specific types of cancer. Similarly, large scale CRISPR functional genomics efforts have mapped genes required for cancer cell proliferation or survival in hundreds of cell lines. Despite this, for specific disease subtypes, such as metastatic prostate cancer, there are likely a number of undiscovered tumor specific driver genes that may represent potential drug targets. To identify such genetic dependencies, we performed genome-scale CRISPRi screens in metastatic prostate cancer models. We then created a pipeline in which we integrated pan-cancer functional genomics data with our metastatic prostate cancer functional and clinical genomics data to identify genes that can drive aggressive prostate cancer phenotypes. Our integrative analysis of these data reveals known prostate cancer specific driver genes, such as *AR* and *HOXB13*, as well as a number of top hits that are poorly characterized. In this study we highlight the strength of an integrated clinical and functional genomics pipeline and focus on two top hit genes, *KIF4A* and *WDR62*. We demonstrate that both *KIF4A* and *WDR62* drive aggressive prostate cancer phenotypes in vitro and in vivo in multiple models, irrespective of AR-status, and are also associated with poor patient outcome.

<sup>1</sup> Department of Radiation Oncology, University of California, San Francisco, San Francisco, CA, USA. <sup>2</sup> Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, USA. <sup>3</sup> Department of Urology, University of California, San Francisco, San Francisco, CA, USA. <sup>4</sup> Division of Hematology and Oncology, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA. <sup>5</sup> Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA. <sup>6</sup> Department of Cellular & Molecular Pharmacology, University of California, San Francisco, San Francisco, CA, USA. <sup>7</sup> These authors contributed equally: Felix Y. Feng, Luke A. Gilbert <sup>8</sup> These authors jointly supervised this work: Felix Y. Feng, Luke A. Gilbert ✉ email: [Felix.Feng@ucsf.edu](mailto:Felix.Feng@ucsf.edu); [Luke.Gilbert@ucsf.edu](mailto:Luke.Gilbert@ucsf.edu)

Prostate cancer is a common cancer and the second leading cause of cancer-related deaths among men in the United States<sup>1</sup>. Androgens are a key driver of prostate cancer cell proliferation, and androgen deprivation therapy (ADT) is the mainstay of treatment for men with metastatic prostate cancer<sup>2,3</sup>. While ADT is initially effective, metastatic prostate cancer patients on ADT will ultimately develop resistance and inevitably progress to lethal metastatic castration-resistant prostate cancer (mCRPC)<sup>4</sup>. There is a critical unmet need to identify new molecular therapeutic targets for patients with mCRPC. We sought to address this issue using both functional genomics and clinical genomics strategies to identify driver genes that are associated with disease progression.

Advances in DNA sequencing have enabled comprehensive genomic analysis of metastatic tumors and have identified adaptive changes in the underlying molecular signaling pathways associated with mCRPC<sup>5–7</sup>. Similarly, the advent of loss-of-function CRISPR functional genomics platforms has systematically revealed which genes are required for cancer cell proliferation and survival<sup>8,9</sup>. Although both of these approaches have generated substantial amounts of data, it remains unclear how best to utilize these strategies to identify driver genes that are specific to the context of a particular disease, such as mCRPC. In principle, the combination of clinical and functional genomics data enables one to distinguish universally essential genes from genes that drive specific cancers that are associated with poor prognosis. We therefore hypothesized that an integrated approach could nominate innovative therapeutic targets for patients with aggressive prostate cancer.

We first performed genome-scale CRISPR-interference (CRISPRi) screens in metastatic prostate cancer models to identify genes required for cellular proliferation or survival. We then created an analytic pipeline that integrated mCRPC functional and clinical genomics data with pan-cancer CRISPR functional genomics data, and in doing so identified a number of previously undescribed prostate cancer-specific driver genes. We demonstrated that two of these genes, *KIF4A* (Kinesin Family Member 4A) and *WDR62* (WD Repeat Domain 62), promote aggressive prostate cancer phenotypes in vitro and in vivo. These experiments, combined with clinical data on these genes, serve to nominate *KIF4A* and *WDR62* as prostate cancer-specific driver genes.

## Results

**Genome-scale CRISPRi screens identify prostate cancer-specific driver genes.** We performed genome-scale screens using our previously described CRISPRi functional genomics platform<sup>10,11</sup> in two mCRPC cell lines, LNCaP and C42B, to identify genes that are required for prostate cancer cell proliferation or survival (Fig. 1A). To begin, we generated multiple malignant and immortalized benign prostate CRISPRi cell-line models that stably express dCas9-BFP-KRAB fusion proteins (Supplementary Fig. 1). LNCaP and C42B cell lines are excellent preclinical in vitro models to study mCRPC<sup>12</sup>. These cell models represent androgen-sensitive (LNCaP) and androgen-insensitive aggressive (C42B) mCRPC. Genome-scale pooled genetic screens were performed by transducing LNCaP or C42B cells stably expressing dCas9-KRAB (hereafter denoted as LNCaPi and C42Bi) with a human genome-scale CRISPRi library<sup>10,11</sup> (Fig. 1A and Supplementary Data 1). Samples were collected at time zero (T0) and after 8 population doublings (T8) (Fig. 1A). Each screen was performed in duplicate. We then used next-generation sequencing to quantify the abundance of the sgRNAs in each population of cells. Results obtained from the replicate screens were highly correlated (Fig. 1B and Supplementary Fig. 2A, B).

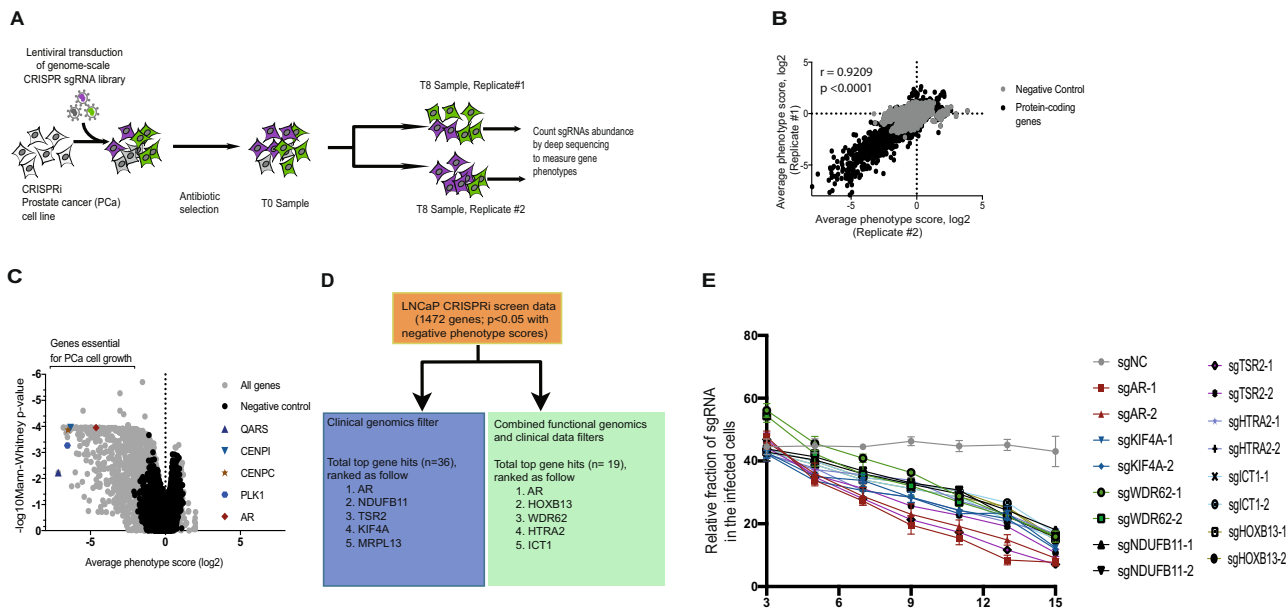
The LNCaP and C42B CRISPRi screens had many shared and selective genetic dependencies (Supplementary Fig. 2C). We were primarily interested in androgen-sensitive mCRPC and so prioritized the LNCaPi screen data. Analysis of our LNCaPi screen data revealed 1472 genes that are required for cell proliferation or survival (Fig. 1C and Supplementary Data 2). Given that the vast majority of these genes were expected to be generally required for cell proliferation, rather than being specifically essential in prostate cancer, we next developed a bioinformatic strategy to identify prostate cancer-specific driver genes.

In order to identify driver genes that are specific to aggressive prostate cancer, we applied a clinical genomics filter designed to prioritize CRISPRi hits with evidence of genomic amplification and/or increased gene expression. This analysis strategy revealed the Androgen Receptor (AR) as the top prostate cancer-specific gene hit, but the next four top-ranked hits were genes not previously associated with prostate cancer: *KIF4A*, *MRPL13*, *NDUFB11*, and *TSR2* (top 5) (Fig. 1D and Supplementary Data 3). Examination of pan-cancer functional genomics data<sup>13,14</sup> for essentiality phenotypes for the top 5 hits revealed that AR is selectively essential for prostate cancer models as expected (Supplementary Fig. 3A). In contrast, top hits such as *TSR2* are essential for the proliferation or survival of nearly all cell types and thus are almost certainly not prostate cancer-specific driver genes (Supplementary Fig. 3B). Other top hits such as *KIF4A*, *MRPL13*, and *NDUFB11* are essential for a number of cell types but not for others (Supplementary Fig. 3C–E). These results demonstrated that solely using clinical data to filter loss-of-function CRISPR functional genomics data can identify known driver genes but also generates a significant rate of false-positive hits as exemplified by *TSR2* in this analysis.

Given these results, we next tested whether a combined clinical and functional genomics filtering strategy would more robustly reveal prostate cancer-specific driver genes in our LNCaPi screen data. We first filtered our list of 1472 LNCaP essential genes against a CRISPR pan-cancer functional genomics datasets (pooled in vitro CRISPR knockout library essentiality screens (PICKLES) and DepMap datasets)<sup>13,15</sup>. We then filtered the remaining hits against two published non-prostate cancer CRISPRi screens<sup>16</sup> to remove CRISPR DepMap false negatives in the prostate cancer CRISPRi screen data. Lastly, we prioritized genes associated with increased expression in metastatic prostate cancer samples (Fig. 1D and Supplementary Data 4). This integrative functional genomics and clinical genomics filtering strategy revealed known prostate cancer-specific driver genes, *AR* and *HOXB13*, as the top two hits<sup>5,7,17,18</sup>. Additional top hits, such as *WDR62*, are uncharacterized in prostate cancer (Fig. 1D). These results demonstrate that an integrated functional genomics and clinical genomics filtering strategy can identify known driver genes and also reveal uncharacterized genes that may drive metastatic prostate cancer progression.

To validate our screen results, we demonstrated that eight top hits from this screen are required for prostate cancer cell proliferation or survival, suggesting our screen results and subsequent analysis nominate reproducible hit genes with a low false-positive rate (Fig. 1E). In order to demonstrate that the hits identified by these two integrated functional and clinical genomics approaches are potential prostate cancer-specific driver genes, we chose to study *KIF4A* and *WDR62* as they were relatively uncharacterized in the context of mCRPC.

***KIF4A* is an AR-independent driver gene in metastatic prostate cancer.** We chose to further investigate *KIF4A* because two independent mCRPC clinical genomics datasets (Quigley, et al.<sup>5</sup> and Abida, et al.<sup>6</sup>) demonstrate that *KIF4A* is a copy number



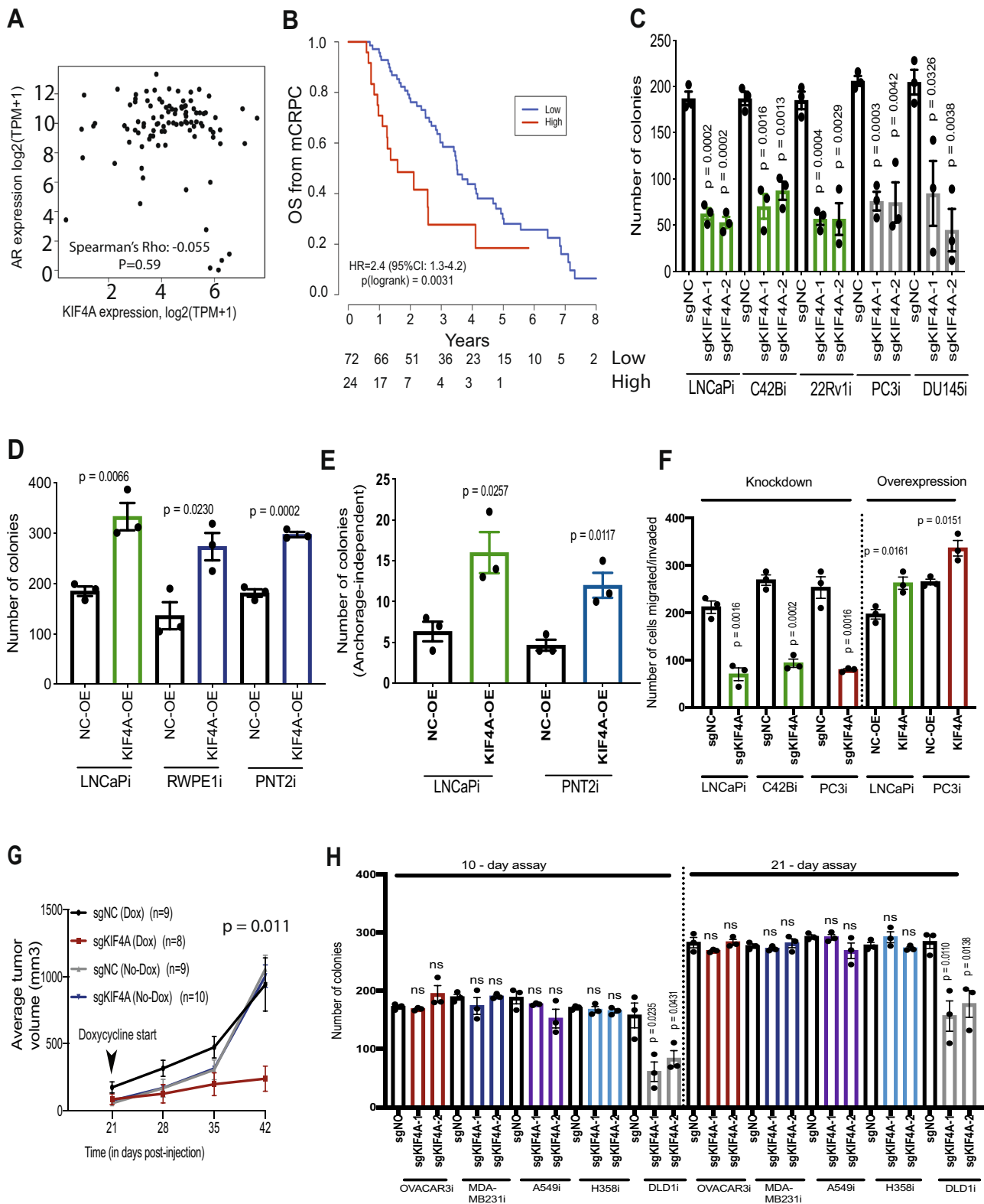
**Fig. 1** Genome-scale CRISPRi screen to identify prostate cancer-specific driver genes. **A** Schematic of growth-based CRISPRi screens; **B** Correlation plot showing concordance between replicates of the CRISPRi screen in LNCaP cells (LNCaPi) at gene level; **C** Volcano plot showing Mann-Whitney statistical significance and average phenotype score of all genes and negative controls; **D** Filtering strategies implemented to identify genes that are specific to prostate cancer biology in the context of metastasis; **E** Competitive growth-based assay to validate top eight hits ( $n = 3$  as biological replicates; Mean  $\pm$  SEM).

amplified in ~33% of mCRPC patient samples (Supplementary Fig. 4A, B) supporting the hypothesis that this may be a mCRPC driver gene. *KIF4A* and *AR* are both on the same arm of the X-chromosome and are frequently co-amplified despite being 2.55 megabases apart<sup>5</sup>. *AR* is also focally amplified in mCRPC<sup>5</sup>. The occurrence of recurrent long amplicons encompassing more than just *AR* supports the hypothesis that *KIF4A*, *AR*, and possibly additional genes may be a ‘cluster’ of cancer driver genes that are co-amplified on the X-chromosome to drive prostate cancer tumorigenesis. In localized primary human prostate cancer samples from the TCGA cohort, we observed that *KIF4A* and *AR* gene expression are correlated (Supplementary Fig. 5A). Surprisingly, in our mCRPC data<sup>5</sup>, we found no correlation between *AR* and *KIF4A* expression, suggesting an independent role for *KIF4A* from *AR* (Fig. 2A). In support of this finding, we also found no correlation between *AR* and *KIF4A* expression in another independent mCRPC cohort (Abida, et al.<sup>6</sup>) (Supplementary Fig. 5B). Importantly, we found that high *KIF4A* expression was associated with poor outcome in mCRPC patients (Fig. 2B). There was also a strong positive correlation between expression of *KIF4A* and of *MKI67*, a marker of cell proliferation in mCRPC samples (in both Quigley et al. and Abida et al.) suggesting *KIF4A* is associated with cancer cell proliferation or survival in patients (Supplementary Fig. 6A, B). Collectively, these data suggest that *KIF4A* may have different functions in primary prostate cancer and mCRPC.

To experimentally dissect the role of *KIF4A* in prostate cancer, we tested whether *KIF4A* expression was required for prostate cancer cell proliferation and/or survival in both AR-dependent (LNCaPi, C42Bi, and 22Rv1i) and AR-independent (PC3i and DU145i) cell-line models. We observed that inhibition of *KIF4A* expression reduced clonogenic survival in both AR-dependent and AR-independent prostate cancer models (Fig. 2C). *KIF4A* knockdown in LNCaPi cells did not decrease AR mRNA or protein levels (Supplementary Fig. 7A, B) *KIF4A* knockdown was similar across this panel of cell lines (Supplementary Fig. 8A). We also observed that *KIF4A* overexpression increased clonogenic

survival in both malignant (LNCaPi) and benign prostate cells (RWPE1i and PNT2i) (Fig. 2D). *KIF4A* overexpression also increased anchorage-independent colony formation in both malignant and benign prostate cells (Fig. 2E). The level of *KIF4A* overexpression was similar across the cell-line models at mRNA level (Supplementary Fig. 8B) and at protein level (Supplementary Fig. 8C). Next, we modulated *KIF4A* expression and performed migration/invasion assays, and observed that the repression of *KIF4A* decreases cell migration/invasion while overexpression of *KIF4A* increases cell migration/invasion in AR-dependent as well as AR-independent cell models (Fig. 2F and Supplementary Fig. 8D). We also performed cell-cycle analysis on LNCaPi and PC3i cells, which are AR-dependent and -independent, respectively, with and without knockdown of *KIF4A*. *KIF4A* knockdown resulted in accumulation of cells in S phase (Supplementary Fig. 9A–D). Lastly, we created an inducible CRISPRi LNCaP (LNCaPi-Dox) model (Supplementary Figs. 10 and 11A). We implanted LNCaPi-Dox cells expressing negative control sgRNAs or sgRNAs targeting *KIF4A* cells into mice. At tumor onset, mice were treated with doxycycline to induce *KIF4A* knockdown or control. We observed that *KIF4A* is necessary for tumor growth in vivo (Fig. 2G). We collected tumor samples 21 days after addition of doxycycline and observed that *KIF4A* repression remained stable at this time point (Supplementary Fig. 11B).

Our clinical filtered screen results nominated *KIF4A* as a prostate-specific driver gene; however, DepMap CRISPR functional genomics data reported that *KIF4A* was essential in a number of cancer cell lines (Supplementary Fig. 3C). To test the model that *KIF4A* is a prostate-specific driver gene, we generated a panel of diverse CRISPRi non-prostate cancer cell-line models (breast: MDA-MB-231i, ovarian: OVCAR3i, lung: H358i and A549i, colon: DLD1i) (Supplementary Fig. 12A). We repressed *KIF4A* in each of these models and then measured the impact of *KIF4A* knockdown on cell proliferation and survival in 10-day and 21-day clonogenic survival assays. In five out of the six cell models, *KIF4A* was dispensable for cell proliferation or survival (Fig. 2H). Similarly, *KIF4A* knockdown in these non-prostate



cancer cells did not show an effect on cell cycle (Supplementary Fig. 12C–H). *KIF4A* knockdown was similar across this panel of cell lines and also similar to *KIF4A* knockdown in prostate cancer cells (Supplementary Figs. 8A and 12B). Basal *KIF4A* protein levels do not correlate with gene essentiality across prostate and non-prostate cell-line models (Supplementary Fig. 13). This suggested that the driver gene properties of *KIF4A* could be relatively specific to prostate biology. Alternately, this discrepancy

could be due to differences in the level of *KIF4A* disruption by CRISPR and by CRISPRi which would suggest that the exact level of *KIF4A* required for cell viability varies between cell types. If this is the case, then our data would demonstrate that there is a therapeutic window for inhibiting *KIF4A* as an anti-cancer strategy.

To further investigate why *KIF4A* is required for prostate cancer proliferation or survival, we measured the transcriptional

**Fig. 2** *KIF4A* is an AR-independent driver gene in metastatic prostate cancer. **A** Scatter plot showing no correlation between *AR* and *KIF4A* in 99 mCRPC patients based on a two-sided Spearman's correlation test (Quigley et al); **B** A Kaplan-Meier curve of overall survival of 96 patients with CRPC with high and low level of *KIF4A*. Differences between groups were tested with a two-sided log-rank test. Hazard ratios were calculated using the Cox proportional hazards regression model. Number at risk is shown under the plot; **C** Colony formation assay in range of prostate cancer cell-line models with *KIF4A* knockdown ( $n = 3$  as biological replicates; Mean  $\pm$  SEM; Unpaired two-tailed *t*-test was used to determine statistical significance); **D** Colony formation assay in malignant and benign prostate cells with *KIF4A* overexpression ( $n = 3$  as biological replicates; Mean  $\pm$  SEM; Unpaired two-tailed *t*-test was used to determine statistical significance); **E** Anchorage-independent growth assay in malignant and benign prostate cells with *KIF4A* overexpression ( $n = 3$  as biological replicates; Mean  $\pm$  SEM; Unpaired two-tailed *t*-test was used to determine statistical significance); **F** Migration and Invasion assay with *KIF4A* knockdown and overexpression in malignant prostate cells ( $n = 3$  as biological replicates; Mean  $\pm$  SEM; Unpaired two-tailed *t*-test was used to determine statistical significance); **G** Line plot showing average tumor volume in *KIF4A* knockdown and control cells implanted in vivo. Average tumor volume was plotted and two-way ANOVA was used to measure statistical significance; **H** Colony formation assay in a range of non-prostate cancer CRISPRi cell-line models with *KIF4A* knockdown ( $n = 3$  as biological replicates; Mean  $\pm$  SEM; Unpaired two-tailed *t*-test was used to determine statistical significance).

consequences of *KIF4A* repression in LNCaPi cells. Using gene set enrichment analysis (GSEA), we observed a signature of multiple cancer signaling pathways, including *MYC* and *E2F*, known to drive cell proliferation (Supplementary Fig. 14A–E). *KIF4A* is a chromokinesin with two reported functions. *KIF4A* is reported to regulate multiple aspects of spindle organization and chromosome positioning/integrity, processes which are thought to be required for all dividing cells. *KIF4A* is also reported to form a complex with both DNMT3B<sup>19</sup>, an enzyme that catalyzes DNA methylation as well as key genes that regulate chromatin, such as HDAC1 and SIN3A<sup>19</sup>, suggesting *KIF4A* could be required for epigenetic programs that support mCRPC. To test whether repression of *KIF4A* alters the chromatin landscape of prostate cancer cells, we performed an Assay for Transposase-Accessible Chromatin using sequencing (ATAC-Seq) experiment in LNCaPi and C42Bi cells. Our data demonstrated that the patterns of open and closed chromatin are broadly remodeled upon *KIF4A* knockdown suggesting *KIF4A* may play a role in the regulation of chromatin biology in prostate cancer cells (Supplementary Figs. 15A–E and 16A, C). The altered ATAC-seq peaks were enriched for pathways which have been established to play crucial role in cancer cells proliferation and survival, including prostate cancer (Supplementary Fig. 16B, D).

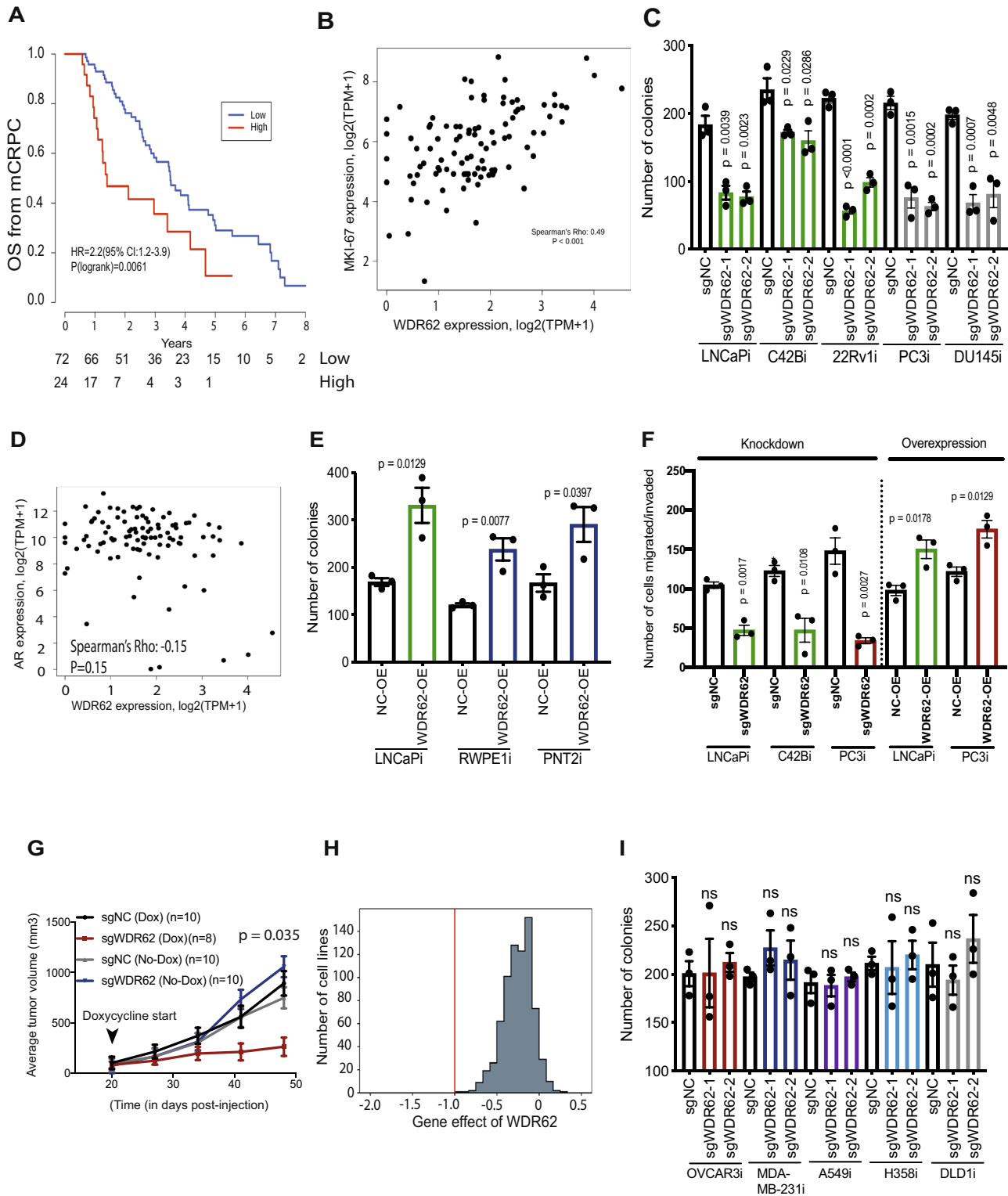
**WDR62 is an uncharacterized prostate cancer driver gene.** We were intrigued by the observation that *WDR62* is an uncharacterized top CRISPRi screen hit gene that is crucial for prostate cancer cell proliferation or survival. In clinical genomics data, we observed that the expression of *WDR62* is significantly higher in primary (TCGA) and metastatic (MSKCC) prostate cancer samples relative to benign prostate samples (Supplementary Fig. 17A, B). We found that high *WDR62* expression is associated with poor patient outcome in mCRPC patients (Quigley et al.) (Fig. 3A). There is also a strong positive correlation between *WDR62* expression and *MKI67*, a marker of cell proliferation in mCRPC samples suggesting *WDR62* could drive tumor cell proliferation or survival (Fig. 3B). These data demonstrate that *WDR62* expression is correlated with disease progression and/or poor outcomes in mCRPC patients.

To experimentally dissect the role of *WDR62* in prostate cancer, we performed clonogenic survival assays in five prostate cancer cell models and observed a significant decrease in the colony formation following knockdown of *WDR62* in all five models demonstrating *WDR62* is required for prostate cancer proliferation or survival irrespective of AR status (Fig. 3C). *WDR62* basal protein expression and knockdown was similar across this panel of cell lines (Supplementary Fig. 18A, B). This is supported by our clinical data demonstrating *WDR62* levels are not correlated with *AR* levels in metastatic prostate cancer (Fig. 3D). Overexpression of *WDR62* significantly increased clonogenic survival of both malignant and benign prostate cells

suggesting *WDR62* is a driver gene in prostate cancer that is likely independent of AR (Fig. 3E). The level of *WDR62* overexpression was similar across prostate malignant and benign cell-line models (Supplementary Fig. 18C). We also observed that knockdown of *WDR62* suppressed the ability of prostate cancer cells to migrate/invade, while *WDR62* overexpression increased migration and invasion (Fig. 3F). Lastly, we subcutaneously implanted LNCaPi-Dox cells expressing control sgRNAs or sgRNAs targeting *WDR62* into mice (Supplementary Fig. 19A). At tumor onset we treated the mice with doxycycline to induce *WDR62* knockdown and observed that *WDR62* is required for tumor growth in vivo (Fig. 3G). Thirty days after addition of doxycycline, we collected the tumors and observed that *WDR62* repression remains stable at that time point (Supplementary Fig. 19B). These results demonstrate *WDR62* is a driver of aggressive prostate cancer phenotypes both in vitro and in vivo.

Our functional genomics filters nominated *WDR62* as being selectively essential in prostate cancer models (Fig. 3H). However, to experimentally confirm this, we repressed *WDR62* expression in our panel of diverse CRISPRi non-prostate cancer cell-line models. In six out of six models, we found that *WDR62* is not required for cell proliferation and survival in a 10-day clonogenic survival assay or progression through the cell cycle (Fig. 3I, Supplementary Fig. 20A–D). In this experiment, *WDR62* was repressed to an equivalent extent in each model as observed in CRISPRi prostate cell lines (Supplementary Fig. 20E). Collectively, these data suggest that *WDR62* is selectively essential in prostate cancer models. Basal protein expression of *WDR62* was similar across this panel of non-prostate cancer cell lines (Supplementary Fig. 20F).

**WDR62 mediates the stability of the TPX2/AURKA protein complex in prostate cancer.** Relatively little is known about *WDR62*<sup>20–25</sup>. Mutations in *WDR62* are associated with microcephaly in humans<sup>20,22,23</sup>. *WDR62* has been reported to interact with Aurora A and TPX2<sup>22,26</sup>. However, this interaction is not annotated in protein–protein interaction databases<sup>27</sup>. In our mCRPC clinical data, we observed a strong positive correlation between *WDR62* and *TPX2* expression levels (Fig. 4A), as well as between *WDR62* and *AURKA* expression levels (Fig. 4B), which can reflect a functional relationship<sup>28</sup>. A similar correlation was also found in an independent mCRPC cohort (Abida et al.<sup>6</sup>) (Supplementary Fig. 21A, B). TPX2 and AURKA bind directly in a known protein complex that is required for mitosis in most cell types and as such both are common essential genes (Supplementary Fig. 21C, D)<sup>29–31</sup>. To test whether AURKA, TPX2, and *WDR62* form a protein complex in mCRPC, we immunoprecipitated *WDR62* from LNCaP cells and then western blotted for TPX2 and AURKA (Fig. 4C). This experiment demonstrated that *WDR62* interacts with TPX2 and AURKA likely forming a protein complex in prostate cancer cells. We hypothesized that



WDR62 may regulate the stability or function of the TPX2/AURKA protein complex in prostate cancer cells. We observed that the knockdown of WDR62 results in loss of TPX2 and AURKA protein suggesting WDR62 regulates the stability of this protein complex in prostate cancer cells (Fig. 4D). Mechanistically, we observed that upon knockdown of WDR62, AURKA is efficiently degraded by the proteasome (Fig. 4E).

To further confirm that WDR62 is a prostate cancer-specific regulator of the TPX2/AURKA protein complex, we examined

WDR62, AURKA, TPX2 genetic dependencies in the DepMap. As expected, TPX2 and AURKA genetic dependencies are highly correlated (Fig. 4F). In contrast, there is no correlation between WDR62 and TPX2 phenotypes in DepMap data compatible with the hypothesis that WDR62 specifically regulates TPX2/AURKA in prostate cancer cells (Fig. 4G). Notably, only a single prostate cancer cell line (VCaP) is represented in the DepMap data and so we do not expect a pan-cancer correlation between TPX2 and WDR62 if the biology of WDR62 is prostate cancer-specific.

**Fig. 3** *WDR62* is an uncharacterized prostate cancer driver gene. **A** A Kaplan–Meier curve of overall survival of 96 patients with CRPC with high and low expression of *WDR62*<sup>5</sup>. Differences between groups were tested with a two-sided log-rank test. Hazard ratios were calculated using the Cox proportional hazards regression model. Number at risk is shown under the plot; **B** Scatter plot showing correlation between expression level of *MKI67* and *WDR62* in 99 mCRPC patients<sup>5</sup>. Spearman’s correlation with a two-sided test for significance was calculated; **C** Colony formation assay in a range of prostate cancer cell-line models with *WDR62* knockdown ( $n = 3$  as biological replicates; Mean  $\pm$  SEM; Unpaired two-tailed  $t$ -test was used to determine statistical significance); **D** Scatter plot ( $n = 99$ ) showing no correlation between *WDR62* and *AR*<sup>5</sup>. Spearman’s correlation with two-sided test for significance was calculated; **E** Colony formation assay in malignant and benign prostate cells with *WDR62* overexpression ( $n = 3$  as biological replicates; Mean  $\pm$  SEM; Unpaired two-tailed  $t$ -test was used to determine statistical significance); **F** Migration and Invasion assay with *WDR62* knockdown and overexpression in malignant prostate cells ( $n = 3$  as biological replicates; Mean  $\pm$  SEM; Unpaired two-tailed  $t$ -test was used to determine statistical significance); **G** Line plot showing average tumor volume in *WDR62* knockdown and control cells implanted in vivo. Average tumor volume was plotted and two-way ANOVA was used to measure statistical significance; **H** Histogram of pan-cancer essentiality CERES scores of *WDR62* in DepMap database. The red line denotes the median gene effect of all common essential genes.; **I** Colony formation assay in a range of non-prostate cancer CRISPRi cell-line models with *WDR62* knockdown ( $n = 3$  as biological replicates; Mean  $\pm$  SEM; Unpaired two-tailed  $t$ -test was used to determine statistical significance).

Together, these data suggest that *WDR62* is a prostate cancer-specific driver gene and in addition is a potential therapeutic target in mCRPC.

## Discussion

Most advanced cancers are driven by the biochemical activity of multiple driver genes. In prostate cancer, AR is the major driver of disease progression, but it remains unclear how additional genes drive mCRPC through AR-dependent and independent mechanisms. In addition, an emerging theme in mCRPC is that not all tumors are addicted to AR<sup>32–34</sup>. Using a genome-scale CRISPRi functional genomics platform, we have identified genes required for proliferation or survival in two models of advanced prostate cancer. Integrative analysis of these data with clinical genomics data and functional genomics data revealed AR as a top hit in these screens but also nominated a number of additional poorly characterized prostate-specific driver genes. Our work highlights the strength of an integrated clinical and functional genomics pipeline and focuses on two of the poorly characterized hit genes. Our data demonstrated *KIF4A* and *WDR62* are AR-independent prostate cancer driver genes that are associated with poor prognosis in patients with advanced metastatic disease.

Our interest in *KIF4A* was driven by the clinical data analysis showing this gene has many of the properties of a human cancer driver gene. In support of this, our data in vitro and in vivo data demonstrated that *KIF4A* promotes phenotypes associated with poor prognosis in mCRPC, irrespective of AR status, suggesting *KIF4A* is a driver gene that promotes disease progression in human prostate cancers. Mechanistically, *KIF4A* biology is complex and the role of *KIF4A* in prostate cancer is fairly unknown. We demonstrated through transcriptome profiling and genome-wide chromatin accessibility analysis that knockdown of *KIF4A* reveals gene signatures that are established to have profound effect on prostate cancer phenotypes. However, open questions remain with respect to how *KIF4A* interacts with other proteins to modulate gene expression, chromatin biology, and prostate cancer phenotypes. We note that *KIF4A* has been nominated as a clinical biomarker of aggressive prostate cancer<sup>35</sup>. Our data currently does not support the hypothesis that *KIF4A* and AR form an auto-regulatory positive feedback loop in prostate cancer<sup>36</sup>. However, it will be important to model *KIF4A* in genetically engineered mouse models of prostate cancer to further explore the biology of how *KIF4A* promotes tumorigenesis, metastasis, and drug response.

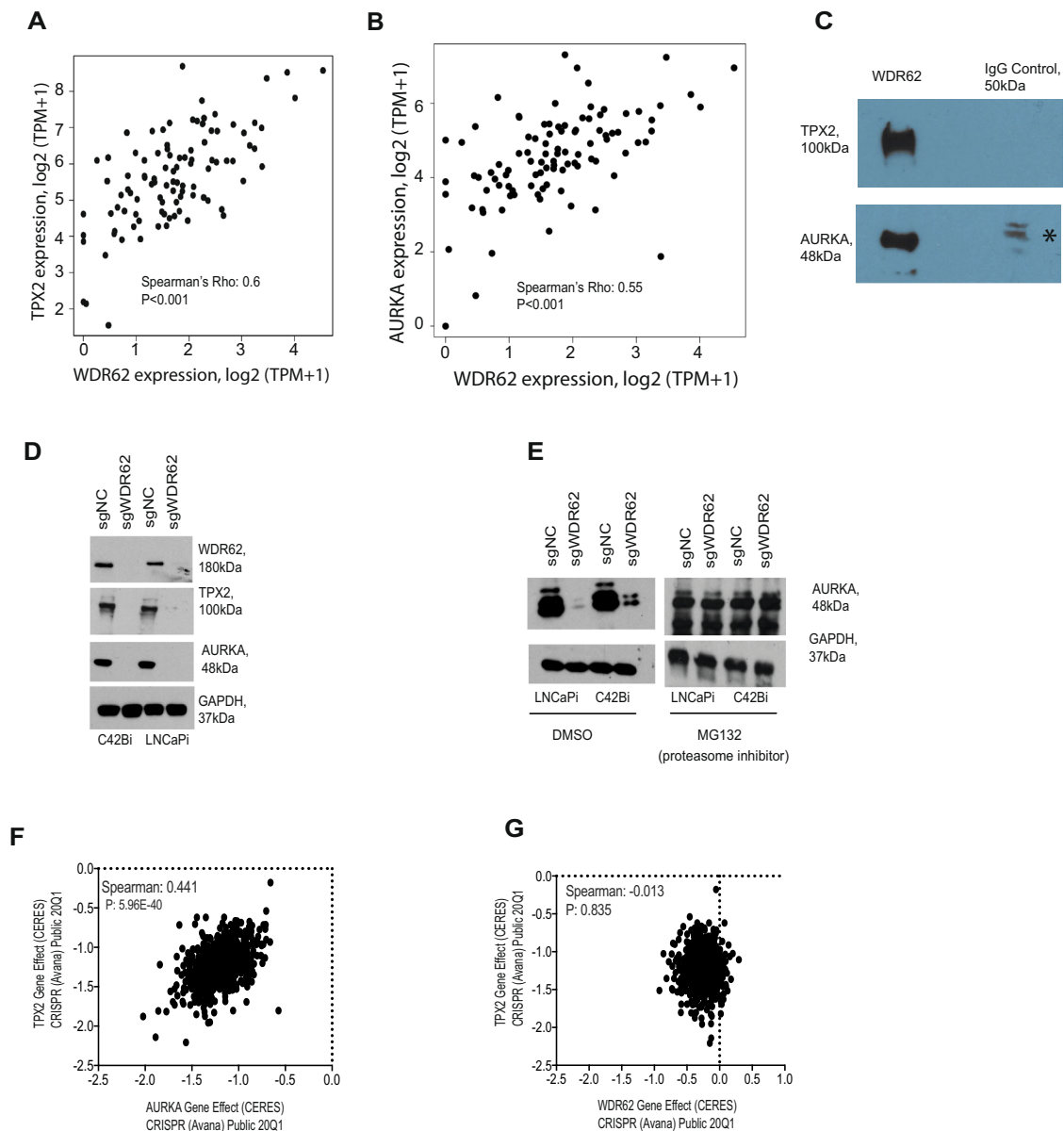
A top hit in our CRISPRi screen is *WDR62*, an uncharacterized gene in prostate cancer. The identification of *WDR62* and additional hit genes as poorly characterized prostate cancer driver genes highlights the strength of our integrated clinical and functional genomics analysis strategy. *WDR62* is one of the most commonly mutated autosomal recessive primary microcephaly-

associated genes, with over 30 identified mutations leading to reduced brain size and a spectrum of cortical abnormalities<sup>20,37</sup>. In this study, we demonstrated that *WDR62* is highly expressed in primary as well as metastatic prostate cancer and is associated with poor patient outcome. We demonstrated both in vitro and in vivo that *WDR62* promotes aggressive prostate cancer phenotypes in all of the prostate cancer models that were tested, irrespective of AR-status, suggesting that this is an AR-independent human prostate cancer driver gene. We note that *WDR62* was previously characterized as a biomarker of resistance to AR targeted therapies in models of neuroendocrine prostate cancer (NEPC) and in NEPC patient samples however to our knowledge *WDR62* has not been previously implicated as a driver of prostate cancer adenocarcinoma<sup>38</sup>. Mechanistically, our data demonstrated that *WDR62* modulates the activity of AURKA by stabilizing the AURKA/TPX2 protein complex in prostate cancer cells. Genetically engineered mouse models of prostate cancer are needed to further explore the biology of how *WDR62* promotes tumorigenesis, metastasis, and drug response. *AURKA* and *TPX2* are required for proliferation in most cells; however, our data and data from DepMap demonstrated that *WDR62* is uniquely essential for proliferation in prostate cancer cells, suggesting *WDR62* could be a potential molecular target in mCRPC patients to disrupt *AURKA-TPX2* signaling without being lethal to other non-prostate tissues<sup>39–41</sup>. However, additional studies are required to prove that *WDR62* does not regulate AURKA/TPX2 activity across other cell types.

To our knowledge, no previous studies have utilized the strengths of integrating clinical genomics and functional genomics data to identify prostate cancer-specific driver genes. Our manuscript leverages an integrated clinical and functional genomics pipeline to identify and validate genes that can drive aggressive prostate cancer phenotypes. As expected, AR was the top hit in our analysis. Development of novel agents targeting mCRPC drivers, other than AR, such as *KIF4A* and *WDR62* may provide important therapeutic strategies for mCRPC patients. Furthermore, for patients with metastatic AR-dependent prostate cancer innovative combination therapies that co-target multiple driver genes simultaneously may lead to increased patient survival.

## Methods

**Cell culture.** LNCaP, C42B, 22Rv1, PC3, DU145, RWPE-1, OVCAR3, MDA-MB-231, A549, H358, DLD1 cell lines were purchased from ATCC. PNT2 cells were purchased from Millipore Sigma. RPMI 1640 medium (Gibco Catalog number: 11875119) + 10% FBS (fetal bovine serum) were used to grow all the cell lines, except RWPE-1 and HEK293T cells. RWPE-1 cells were grown using Keratinocyte Serum Free Medium (Gibco Catalog number: 17005042) + 10% FBS. HEK293T cells were grown using DMEM medium (Gibco Catalog number: 11885084) + 10% FBS. All cell lines underwent verification by short tandem repeat profiling at Genetica Cell-Line Testing.



**Fig. 4** WDR62 mediates the stability of the TPX2/AURKA protein complex in prostate cancer. **A** and **B** Scatter plots ( $n = 99$ ) showing a correlation between TPX2 and WDR62 and AURKA and WDR62, respectively<sup>5</sup>. Spearman's correlation with a two-tailed test for significance was calculated; **C** Co-immunoprecipitation of WDR62 with TPX2 and AURKA (\* non-specific band). The co-immunoprecipitation experiment was performed twice to determine reproducibility; **D** Western blot showing loss of AURKA and TPX2 following knockdown of WDR62. Each western blot experiment was performed twice to determine reproducibility; **E** Western blot of AURKA following knockdown of WDR62 with and without MG132, a proteasome inhibitor. Each western blot experiment was performed twice to determine reproducibility; **F** and **G** Scatter plots showing phenotype (gene effect) correlation between TPX2 and AURKA and TPX2 and WDR62, respectively<sup>14</sup>. Spearman's correlation was performed for statistical analysis.

**CRISPR vectors.** Lentiviral vectors were used to express the CRISPRi protein and sgRNAs in human cells as previously described<sup>10</sup>. The CRISPRi protein (dCas9-BFP-KRAB) was expressed from either the SFFV or TET3G promoter. The sgRNA vector encodes an sgRNA driven by the mouse U6 promoter as well as a fluorescent protein (BFP or GFP) T2A puromycin N-acetyl transferase gene driven by the human EF1 $\alpha$  promoter. The sgRNA sequences used in experiments are provided in Supplementary Table 2.

**Lentivirus generation:** HEK293T cells were used to generate lentivirus. pCMV-dR8.91 and pMD2-G were used as packing vectors. LT1 transfection reagent (Mirus MIR2300) was used for transfection. To enhance viral production viral boost reagent (Alstem # VB100) was used.

**Cell-line construction.** For constitutive and inducible CRISPRi cell lines, polyclonal cells expressing dCas9-BFP-KRAB fusion proteins driven from an SFFV or TRE3G promoter, respectively, were generated by viral transduction followed by fluorescence-activated cell sorting to purity using a BD Fusion. Prostate cancer CRISPRi lines are denoted as LNCaPi, C42Bi, 22Rv1i, PC3i, and DU145i. LNCaPi,

C42Bi, and 22Rv1i AR-dependent prostate cancer lines, whole PC3i and DU145i are AR-independent cell lines. Benign prostate CRISPRi lines are denoted as RWPE1i and PNT2i. Doxycycline inducible CRISPRi line on LNCaPi line is denoted as LNCaPi-Dox. Non-prostate cancer CRISPRi lines are denoted as OVCAR3i (is an ovarian cancer line), MDA-MB-231i (is a breast cancer line), A549i and H358i (are lung cancer lines), and DLD1i (is a colon cancer line).

**CRISPRi screen.** The human CRISPRi V2 Top5 sgRNA library was used to perform CRISPRi screens<sup>11</sup>. This sgRNA library targets 18,905 human genes with five sgRNAs per transcription start site. Cells were grown at minimum library coverage of 1000x for genome-scale screens. Cells were collected at 0 and 8 doubling after puromycin selection and harvested cells were processed for next-generation sequencing. The screens were performed in two technical replicates. Briefly, DNA was isolated, the cassette encoding the sgRNA was amplified by PCR, and relative sgRNA abundance was determined by next-generation sequencing as previously described<sup>10,11</sup>. Data were analyzed using publicly available code (<https://github.com/mhorlbeck/ScreenProcessing>).



**Clinical cohorts.** Quigley et al.<sup>5</sup> is the published whole-genome and -transcriptome analysis of 101 castration-resistant prostate cancer metastases. Gene expression and copy number data were processed as described in the original publication. Abida et al.<sup>6</sup> TCGA<sup>42</sup> and MSKCC<sup>43</sup> datasets were downloaded through the cBioPortal for Cancer Genomics<sup>44,45</sup>, and gene expression and copy number data were analyzed as processed by cBioPortal. All analyses on these cohorts were performed in the R statistical environment version 3.6.3 using RStudio 1.2.503<sup>46,47</sup>. The correlation was calculated by Spearman's rank correlation. A two-sided Wilcoxon rank-sum test was used to test for differences between two groups, unless otherwise stated. Violin plots were created using the ggplot2 package<sup>48</sup>. Survival analyses for the Quigley et al. study were performed with overall survival as endpoint, calculated from the time of mCRPC diagnosis until death or censoring, and visualized using the Kaplan–Meier method using the survival package and plotting functions from the rms package<sup>49–52</sup>. Differences between groups were tested with a two-sided log-rank test. Hazard ratios were calculated using the Cox proportional hazards regression model. High expression was defined as the top quartile of gene expression for the respective genes.

**Clinical genomics filters.** Considering that drivers of mCRPC commonly have copy number gains<sup>6,7</sup>, first, we ranked the genes ( $n = 1472$ ) by number of mCRPC samples having a copy number gain of the gene in the Quigley et al. We then further narrowed down this list of genes by the criteria that a copy number gain would result in a corresponding change in phenotype (gene expression) due to the copy number gain. The cut-off for the change in gene expression with or without copy number gain was set to be  $\geq 2$  fold change in mean expression (transcripts per million, TPM) with  $p$  value  $\leq 0.001$  (two-sided Wilcoxon rank-sum test).

**Combined functional genomics and clinical data filters.** The goal of this filter was to identify prostate cancer-specific genes that are expressed higher in metastases compared to primary prostate tumors and/or benign prostate. First, the 1472 genes were filtered against the published pan-cancer, except prostate cancer PICKLES database<sup>15</sup>. Genes that were called as essential in 90% of the cell lines were removed from the gene list. Second, the gene list was filtered through the CERES cancer dependency database (DepMap, Version 20Q1)<sup>13</sup>. Genes called essential by the DepMap database were removed from the list. Third, we further removed any gene on the list that were found to be as a significant hit in two non-prostate cancer CRISPRi screen published recently<sup>16</sup>. Finally, the genes whose expression were found to be high in metastasis samples (in MSKCC cohort) compared to primary tumor were only considered. The narrowed-down gene list was then ranked based on their screen phenotype score and  $p$  value (smallest to largest).

For DepMap gene effect data, as defined by DepMap a score of 0 is equivalent to a gene that is not essential, whereas a score of  $-1$  corresponds to the median of all common essential genes<sup>14,53,54</sup>.

**Competitive growth assay.** LNCaPi cells plated out at 150,000 cells per well in six-well plates were infected with lentivirus sgRNA targeting gene of interest and control. Each lentiviral transduction of sgRNA and control were done in three biological replicates and performed at least twice to determine the reproducibility of the data. Data were collected and analyzed using Invitrogen Attune NxT Flow Cytometer at day 3 post-infection and every 48 h thereafter. The relative fraction of sgRNA in the infection cells was measured.

**Clonogenic assay.** Cells transduced with the indicated sgRNAs and then were treated with puromycin to achieve a pure population of the transduced cells. One thousand cells were seeded out per well in a six-well plate. Each experimental setup was performed in three biological replicates and done at least thrice to determine the reproducibility of the data. RNA was isolated from the remaining cells to confirm knockdown by qRT-PCR. The cells were allowed to grow for 4 days. On day 5, the wells were washed with PBS, and colonies are fixed with 25% methanol and stained with crystal violet (0.05% w/v). Images were scanned and analyzed using ImageJ software. Average number of colonies formed was counted and two-tailed  $t$ -test was used to determine statistical significance.

**Migration and invasion assay.** Cells transduced with sgRNAs targeting the gene of interest and control were treated with puromycin for a pure population. A total of 100,000 cells were then plated in serum-free media for invasion in Matrigel-coated micropore transwells (Corning Bio-Coat). Each experimental setup was performed in three biological replicates and done at least thrice to determine the reproducibility of the data. RNA was isolated from the remaining cells to confirm knockdown by qRT-PCR. Three days after invasion/migration plating, transwells were fixed in 4% formaldehyde, stained with crystal violet, and cells on the upper side of the micropore membrane wiped off. Cells on the lower side of the membrane were photographed with a microscope, the dye was dissolved in 10% acetic acid and optical density measured at 560 nm. Average number of cells migrated was counted and two-tailed  $t$ -test was used to determine statistical significance.

**Overexpression vectors.** Mammalian gene collection fully sequenced human KIF4A cDNA (Horizon, Cat #MHS6278-202758261) and human WDR62 (Horizon, Cat #MHS6278-202830915) was used to transfect cells and overexpress KIF4A and WDR62, respectively. Overexpression efficiency was measure either by qPCR or by western blotting.

**Soft-agar assay.** Growth in anchorage-independent condition was assessed by colony formation in low melting agarose. Melted agar solution (1%) was plated out in 10 cm plates and allowed to solidify. Cells transduced to have gene overexpression and control were counted and prepared. Each experimental setup was performed in three biological replicates and done at least thrice to determine the reproducibility of the data. The cell suspension was mixed with 0.6% low melting agar (1:1) and plated out over the solidified agar. The cells were allowed to grow for 21 days. The colonies were then stained with nitroblue tetrazolium chloride. Images were scanned and analyzed using ImageJ software. Average number of colonies formed was counted and two-tailed  $t$ -test was used to determine statistical significance.

**Cell-cycle assay.** Stably expressing dCas9-KRAB CRISPRi cells were plated out at 150,000 cells per well and lentiviral transduced with sgRNA targeting KIF4A and control sgRNA. Post puromycin selection, cells were stained with manufacturers protocol<sup>55</sup>. Briefly, harvested cells were washed with PBS, counted, fixed with 70% ethanol and stained with Propidium Iodide Solution (Cat# 421301). Each experimental setup was performed in three biological replicates and done at least thrice to determine the reproducibility of the data.

**In vivo experiments.** Doxycycline inducible LNCaP cells were lentiviral transduced with sgRNA targeting KIF4A (and also WDR62) and control sgRNA. Post puromycin selection, cells were counted, and the cell suspension was mixed with Matrigel (1:1). A total of  $2 \times 10^6$  cells were injected subcutaneously on each flank of the mice. NOD-SCID-Gamma mice were used. UCSF IACUC and laboratory animal resource center (LARC) mouse husbandry standards were followed for all mouse housing (<https://larc.ucsf.edu/>). Once the tumors were palpable the mice were randomly categorized in two groups receiving either doxycycline diet (Bio-Serv, Cat#S3888) or control diet. Tumors were measured using digital caliper by two different laboratory personnel to rule out any measurement bias. Tumor volume was calculated using the equation: Volume = length  $\times$  width<sup>2</sup>  $\times$  0.52, where the length represents the longer axis. Average tumor volume was plotted and two-way ANOVA was used to measure statistical significance. The mice were humanely euthanized once the tumor reached 1000 cubic mm of size following appropriate UCSF's LARC protocol. Protein was extracted from the tumors of the mice to determine knockdown efficiency. All animal experiments conducted were reviewed and approved by UCSF IACUC board.

**RNA extraction and qPCR.** Cells were viral transduced with sgRNA targeting the gene of interest or control sgRNA for 3 days and selected with puromycin for pure population for 3 days. The entire setup of the experiment was done in three biological replicates. Post-selection RNA was extracted from the cells as per the manufacturer's protocol using the Zymo Quick-RNA-extraction kit Cat# R1054. RNA quantification was done using Thermo Scientific Nanodrop 2000 system. cDNA was prepared using SuperScript III First-Strand Synthesis System for RT-PCR (Cat # 18080-051). QuantStudio Flex Real-Time PCR system was used to measure mRNA expression of gene of interest. The list of all primers used is provided in Supplementary Table 1.

**Western blot.** Cells were lysed using RIPA buffer with protease inhibitor (Thermo, Cat#78430), sonicated and centrifuged to extract protein. Either NuPAGE 4–12% Bis-Tris or 3–8%Tris-Acetate precast polyacrylamide gels were used for protein analysis<sup>56,57</sup>. Antibodies used were as follows: KIF4A (Thermo, Cat# PA5-30492), WDR62 (Bethyl Labs, Cat#A301-560A), GAPDH (14C10) (Cell signaling, Cat#2118), Aurora A (D3E4Q) (Cell signaling, Cat #14475), TPX2 (Novus Bio, Cat#NB500-179), AR (D6F11) (Cell Signaling #5153). Antibody specificity for primary antibodies was validated by CRISPRi knockdown experiments. All antibodies were raised for human proteins and used to detect human proteins. All the western blot experiments were at least done twice to determine reproducibility.

**RNA-Seq sample preparation.** LNCaPi cells were viral transduced with sgRNA targeting KIF4A or control sgRNA for 3 days and selected with puromycin for pure population for 3 days. Post-selection RNA was extracted from the cells as described above for library preparation to generate Illumina compatible libraries of sequences and to perform qPCR to determine the sgRNA knockdown efficiency. QuantSeq 3' mRNA-Seq library prep kit FWD for Illumina (Lexogen, Cat# 015.24) was used to prepare the library as per the manufacturer's protocol. Quality control was performed by using the Agilent Bioanalyzer 2100 system and the samples were sequenced in Illumina HiSeq 4000. The entire set of the experiment was done in two biological replicates.

**RNA-seq data processing.** The RNA-seq single-end fastq data generated by Illumina HiSeq 4000 sequencing system were first trimmed to remove adapter sequence using Cutadapt v2.6<sup>58</sup> with “-q 10 -m 20” option. After adapter trimming, FASTQC v0.11.8<sup>59</sup> was used to evaluate the sequence trimming as well as overall sequence quality. Using splice-aware aligner STAR (2.7.1a)<sup>60</sup>, RNA-seq reads were aligned onto the Human reference genome build hg38 using “-outSAMtype BAM SortedByCoordinate-outSAMunmapped Within-outSAMmapqUnique 50-sjdbOverhang 65-chimSegmentMin 12-twopassMode Basic” option and exon-exon junctions, according to the known Human gene model annotation from the GENCODE v30<sup>61</sup>. Apart from protein-coding genes, non-coding RNA types, and pseudogenes are further annotated and classified. Furthermore, based on the reads that can only be mapped to a single genomic location, the transcript/gene expression quantification was performed using “featurecount” function within Rsubread R-package<sup>62</sup> with “GTF.featureType = “exon”, GTF.attrType = “gene\_id”, useMetaFeatures = TRUE, allowMultiOverlap = FALSE, countMultiMappingReads = FALSE, isLongRead = FALSE, ignoreDup = FALSE, strandSpecific = 0, juncCounts = TRUE, genome = NULL, isPairedEnd = FALSE, requireBothEndsMapped = FALSE, checkFragLength = FALSE, countChimericFragments = TRUE, autosort = TRUE” option. Cross-sample normalization of expression values and differential expression analysis between the KIF4A-knockdown and control was done using DESeq2 R-package<sup>63</sup>. Benjamini-Hochberg corrected *p*-value < 0.05 and log<sub>2</sub> fold change >0.5 or <0.5 were considered statistically significant.

The RNA-seq data generated in this study have been deposited in the NCBI’s Gene Expression Omnibus (GEO) database under accession code GSE178330.

**ATAC-Seq sample preparation.** We performed ATAC-Seq on LNCaP and C42B cells following knockdown of KIF4A or control. The experiment was carried out as described in the published method papers by Buenrostro, et al<sup>64</sup>. and Corces, et al<sup>65</sup>, with the following modifications. Cells were resuspended in buffer (Illumina Cat#), incubated on ice for 10 min, and lysed using a dounce homogenizer. 50,000 nuclei were incubated with 25uL 2X TD Buffer and 1.25uL Transposase (Illumina Tagment Enzyme/Buffer Cat# 20034210) shaking at 300 rpm at 37 C for 30 min. Zymo DNA Clean and Concentrator 5 kit (Cat# D4014) was then used to purify DNA. Transposed DNA was amplified using PCR master mix and indexes from Nextera DNA Library Prep kit (Cat# 15028211) for 5 cycles and then assessed using qPCR. Final cleanup was performed using 1.8X AMPure XP beads (Cat# A63881) and libraries quantified using the DNA High Sensitivity Agilent 2100 Bioanalyzer System. Samples were sequenced at the UCSF Core Facility on the NovaSeq, paired end. The entire experimental setup was performed in two technical replicates.

**ATAC-seq data processing.** The ATAC-seq paired-end fastq data generated by Illumina NovaSeq 6000 sequencing system were first trimmed to remove Illumina Nextera adapter sequence using Cutadapt v2.6<sup>58</sup> with “-q 10 -m 20” option. After adapter trimming, FASTQC v0.11.8<sup>59</sup> was used to evaluate the sequence trimming as well as overall sequence quality. Bowtie2 version 2.3.5.1<sup>66</sup> was then used to align the ATAC-seq reads against the Human reference genome build hg38 using “-very-sensitive” option. The uniquely mapped reads were obtained in SAM format. Samtools version 1.9<sup>67</sup> was used to convert SAM to BAM file as well as sort the BAM file. Picard (<https://broadinstitute.github.io/picard/>) was then used to remove duplicates using the MarkDuplicates tool using “REMOVE\_DUPLICATES = true” option. The resulting BAM file reads position were then corrected by a constant offset to the read start (“+” stranded +4 bp, “-” stranded -5 bp) using deepTools2 v3.3.2<sup>68</sup> with “alignmentSieve-ATACshift” option. This resulted in the final aligned, de-duplicated BAM file that was used in all downstream analyses. ATAC-seq peak calling was performed using MACS2 v2.2.5<sup>69</sup> to obtain narrow peaks with “callpeak -f BAMPE -g hs-nomodel -B-keep-dup all-call-summits” option. The resulting peaks that map to the mitochondrial genome or genomic regions listed in the ENCODE hg38 blacklist (<https://www.encodeproject.org/annotations/ENCSR636HFF/>) or peaks that extend beyond the ends of chromosomes were filtered out. Non-overlapping unique ATAC-seq narrow peaks regions were obtained from the all samples analyzed. Only those non-overlapping unique peak regions present in at least two samples were considered for further analysis. Sequencing reads mapped to these non-overlapping unique regions were counted using “featurecount” function within Rsubread R-package<sup>62</sup> with “isPairedEnd = -TRUE, countMultiMappingReads = FALSE, maxFragLength = 100, autosort = TRUE” option. Further normalization of the feature counts and differential open chromatin regions between KIF4A-knockdown and control were obtained using DESeq2 R-package<sup>63</sup>. Only those peak regions with Benjamini-Hochberg corrected *p*-value < 0.05 and log<sub>2</sub> foldchange >0.5 or <0.5 were considered statistically significant.

The ATAC-seq data generated in this study have been deposited in the NCBI’s GEO database under accession code GSE178330 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE178330>)

**Gene set enrichment analysis (GSEA).** We used GSEA<sup>70</sup> to identify the signaling pathways enriched in the differentially expressed genes between KIF4A-

knockdown and control obtained from RNA-seq analysis. For this, we used fgsea R-package<sup>71</sup> with Hallmark pathway collection from MSigDB<sup>72</sup>.

**Statistical analysis.** In the Quigley et al. study, Spearman’s correlation was used to determine statistical significance for all the correlation plots: *KIF4A* and *AR*; *MKI67* and *KIF4A*; *WDR62* and *AR*; *MKI67* and *WDR62*. For gene expression and correlation, a two-sided Wilcoxon rank-sum test was used to test for differences between two groups, unless otherwise stated. Survival analyses for the Quigley et al. study were performed with overall survival as endpoint, calculated from time of mCRPC diagnosis until death or censoring, and visualized using the Kaplan–Meier method. Differences between groups were tested with a two-sided log-rank test. Hazard ratios were calculated using the Cox proportional hazards regression model. Spearman’s or Pearson correlation analysis was used for all other datasets: prostate cancer TCGA, Abida et al. and MSKCC.

Unpaired *t*-test was used to determine statistical analysis for all the column plots: colony formation assay, anchorage-independent assay, migration and invasion assay, and qPCR results. Two-way ANOVA was used to determine statistical significance in the in vivo data. Spearman’s correlation analysis was performed a correlation of CERES data between TPX2 and AURKA and TX2 and WDR62. In RNA-Seq data, Benjamini–Hochberg test was performed. Corrected *p* value < 0.05 and log<sub>2</sub> foldchange >0.5 or <0.5 were considered statistically significant. In ATAC-Seq data peak regions with Benjamini–Hochberg corrected *p*-value < 0.05 and log<sub>2</sub> foldchange >0.5 or <0.5 were considered statistically significant.

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

## Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files. The RNA sequencing and ATAC sequencing datasets generated in this study have been deposited in the NCBI’s Gene Expression Omnibus (GEO) database under accession code GSE178330. PICKLES (<http://pickles.hart-lab.org>) and DepMap (<https://depmap.org/portal/>) data are publicly available. A reporting summary for this article is available as a Supplementary Information file.

## Code availability

Custom code used in this manuscript is publicly available at <https://github.com/GilbertLabUCSF/CanDI> or freely available on request.

Received: 8 May 2020; Accepted: 15 July 2021;

Published online: 29 July 2021

## References

- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA Cancer J. Clin.* **69**, 7–34 (2019).
- Massie, C. E. et al. The androgen receptor fuels prostate cancer by regulating central metabolism and biosynthesis. *EMBO J.* **30**, 2719–2733 (2011). 6.
- Meng, M. V. et al. Contemporary patterns of androgen deprivation therapy use for newly diagnosed prostate cancer. *Urology* **60**, 7–11 (2002).
- Kirby, M., Hirst, C. & Crawford, E. D. Characterising the castration-resistant prostate cancer population: a systematic review. *Int. J. Clin. Pract.* **65**, 1180–1192 (2011).
- Quigley, D. A. et al. Genomic hallmarks and structural variation in metastatic prostate cancer. *Cell* **174**, 758–769.e9 (2018). 26.
- Abida, W. et al. Genomic correlates of clinical outcome in advanced prostate cancer. *Proc. Natl Acad. Sci. USA* **116**, 11428–11436 (2019). 4.
- Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015). 21.
- Tyner J. W. Integrating functional genomics to accelerate mechanistic personalized medicine. *Cold Spring Harb. Mol. Case Stud.* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5334473/> (2017).
- Behan, F. M. et al. Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature* **568**, 511–516 (2019).
- Gilbert, L. A. et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014). 23.
- Horlbeck M. A. et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* **5**, e19760 (2016).
- Comparative genomic and transcriptomic analyses of LNCaP and C4-2B prostate cancer cell lines. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0090002> (2020).
- Tsherniak, A. et al. Defining a cancer dependency map. *Cell* **170**, 564–576.e16 (2017). 27.

14. Meyers, R. M. et al. Computational correction of copy-number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
15. Lenoir, W. F., Lim, T. L. & Hart, T. PICKLES: the database of pooled in-vitro CRISPR knockout library essentiality screens. *Nucleic Acids Res.* **46**, D776–D780 (2018). 4.
16. Lou K., et al. KRASG12C inhibition produces a driver-limited state revealing collateral dependencies. *Sci. Signal.* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6871662/> (2019).
17. Ewing, C. M. et al. Germline mutations in HOXB13 and prostate-cancer risk. *N. Engl. J. Med.* **366**, 141–149 (2012). 12.
18. Norris, J. D. et al. The homeodomain protein HOXB13 regulates the cellular response to androgens. *Mol. Cell* **36**, 405–416 (2009). 13.
19. Geiman, T. M. et al. Isolation and characterization of a novel DNA methyltransferase complex linking DNMT3B with components of the mitotic chromosome condensation machinery. *Nucleic Acids Res.* **32**, 2716–2729 (2004).
20. Nicholas, A. K. et al. WDR62 is associated with the spindle pole and is mutated in human microcephaly. *Nat. Genet.* **42**, 1010–1014 (2010).
21. Alshawaf A. J., Antonic A., Skafidas E., Ng D. C.-H., Dottori M. WDR62 regulates early neural and glial progenitor specification of human pluripotent stem cells. *Stem Cells Int.* <https://www.hindawi.com/journals/sci/2017/7848932/> (2017).
22. Chen, J.-F. et al. Microcephaly disease gene Wdr62 regulates mitotic progression of embryonic neural stem cells and brain size. *Nat. Commun.* **5**, 1–13 (2014). 30.
23. Jayaraman, D. et al. Microcephaly proteins Wdr62 and aspm define a mother centriole complex regulating centriole biogenesis, apical complex, and cell fate. *Neuron* **92**, 813–828 (2016). 23.
24. Shohayeb, B. et al. The association of microcephaly protein WDR62 with CPAP/IFT88 is required for cilia formation and neocortical development. *Hum. Mol. Genet.* **29**, 248–263 (2020). 15.
25. Qin Y. et al. WDR62 is involved in spindle assembly by interacting with CEP170 in spermatogenesis. *Development.* <http://dev.biologists.org/content/146/20/dev174128> (2019).
26. Lim, N. R. et al. Opposing roles for JNK and Aurora A in regulating the association of WDR62 with spindle microtubules. *J. Cell Sci.* **128**, 527–540 (2015). 1.
27. Szklarczyk, D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
28. Taggart, J. C. & Li, G.-W. Production of protein-complex components is stoichiometric and lacks general feedback regulation in eukaryotes. *Cell Syst.* **7**, 580–589.e4 (2018). 26.
29. Kufer, T. A. et al. Human TPX2 is required for targeting Aurora-A kinase to the spindle. *J. Cell Biol.* **158**, 617–623 (2002). 19.
30. Fu, J., Bian, M., Liu, J., Jiang, Q. & Zhang, C. A single amino acid change converts Aurora-A into Aurora-B-like kinase in terms of partner specificity and cellular function. *Proc. Natl Acad. Sci. USA* **106**, 6939–6944 (2009). 28.
31. Huttlin, E. L. et al. The BioPlex network: a systematic exploration of the human interactome. *Cell* **162**, 425–440 (2015). 16.
32. Beltran, H. et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat. Med.* **22**, 298–305 (2016).
33. Bluemn, E. G. et al. Androgen receptor pathway-independent prostate cancer is sustained through FGF signaling. *Cancer Cell* **32**, 474–489.e6 (2017). 09.
34. Liu Y., et al. The androgen receptor regulates a druggable translational regulon in advanced prostate cancer. *Sci. Transl. Med.* <https://stm.sciencemag.org/content/11/503/eaaw4993> (2019).
35. Gao, H., Chen, X., Cai, Q., Shang, Z. & Niu, Y. Increased KIF4A expression is a potential prognostic factor in prostate cancer. *Oncol. Lett.* **15**, 7941–7947 (2018).
36. Cao, Q. et al. Targeting the KIF4A/AR axis to reverse endocrine therapy resistance in castration-resistant prostate cancer. *Clin. Cancer Res.* **26**, 1516–1528 (2020). 15.
37. Bilgüvar, K. et al. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* **467**, 207–210 (2010).
38. Lin, D. et al. High fidelity patient-derived xenografts for accelerating prostate cancer discovery and drug development. *Cancer Res.* **74**, 1272–1283 (2014). 15.
39. Beltran, H. et al. A Phase II trial of the Aurora kinase A inhibitor alisertib for patients with castration-resistant and neuroendocrine prostate cancer: efficacy and biomarkers. *Clin. Cancer Res. J. Am. Assoc. Cancer Res.* **25**, 43–51 (2019). 1.
40. Beltran, H. et al. Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer Discov.* **1**, 487–495 (2011).
41. Kivinummi, K. et al. The expression of AURKA is androgen regulated in castration-resistant prostate cancer. *Sci. Rep.* **7**, 17978 (2017). 21.
42. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell.* **163**, 1011–1025 (2015).
43. Taylor, B. S. et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22 (2010). 13.
44. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
45. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
46. RStudio | Open source & professional software for data science teams <https://rstudio.com/> (2021).
47. R: the R project for statistical computing <https://www.r-project.org/> (2021).
48. Welcome | ggplot2 <https://ggplot2-book.org/> (2021).
49. Therneau T. M. Survival analysis [R package survival version 3.2-11]. Comprehensive R Archive Network (CRAN). <https://CRAN.R-project.org/package=survival> (2021).
50. Borgan Ø. Modeling survival data: extending the Cox model. (eds. Therneau a, T. M. & Grambsch, P. M.) (Springer-Verlag, New York, 2000). No. of pages: xiii + 350. ISBN 0-387-98784-3.
51. rms: Regression modeling strategies. Comprehensive R Archive Network (CRAN). <https://CRAN.R-project.org/package=rms> (2021).
52. Ghandi, M. et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
53. Yogodzinski C., Arab A., Pritchard J. R., Goodarzi H., Gilbert L. A. A global cancer data integrator reveals principles of synthetic lethality, sex disparity and immunotherapy. Preprint at *bioRxiv* <https://www.biorxiv.org/content/10.1101/2021.01.08.425918v1.full> (2021).
54. Pacini, C. et al. Integrated cross-study datasets of genetic dependencies in cancer. *Nat. Commun.* **12**(Mar), 1661 (2021). 12.
55. Protocol - propidium iodide cell cycle staining protocol. <https://www.biologend.com/en-us/protocols/propidium-iodide-cell-cycle-staining-protocol> (2020).
56. Das, R. et al. MicroRNA-194 promotes prostate cancer metastasis by inhibiting SOCS2. *Cancer Res.* **77**, 1021–1034 (2017). 15.
57. Paltoglou, S. et al. Novel androgen receptor co-regulator GRHL2 exerts both oncogenic and anti-metastatic functions in prostate cancer. *Cancer Res.* **77**, 3417–3430 (2017). 1.
58. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
59. Andrews S. FastQC: A quality control tool for high throughput sequencing data <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
60. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). 1.
61. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019). 8.
62. Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* **47**, e47 (2019).
63. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
64. Buenrostro J., Wu B., Chang H., Greenleaf W. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015).
65. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**(Oct), 959–962 (2017).
66. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
67. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
68. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
69. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
70. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005). 25.
71. Korotkevich G., Sukhov V., Sergushichev A. Fast gene set enrichment analysis. Preprint at *bioRxiv* <https://www.biorxiv.org/content/10.1101/060012v3> (2019).
72. Liberzon, A. et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015). 23.

## Acknowledgements

L.A.G. is supported by K99/R00 CA204602 and DP2 CA239597 as well as the Goldberg-Benioff Endowed Professorship in Prostate Cancer Translational Biology. F.Y.F. is supported by R01 CA230516 and R01 CA227025 as well as a Prostate Cancer Foundation Challenge Grant and the Benioff Initiative for Prostate Cancer Research. D. A.Q. is supported by a Prostate Cancer Foundation Young Investigator Award. J.C. is supported by a PCF Young Investigator Award and DOD award W81XWH-20-1-

0136. M.S. is supported by the Swedish Research Council (Vetenskapsrådet) with grant number 2018-00382 and the Swedish Society of Medicine (Svenska Läkaresällskapet). A.A. is supported by R01 CA230516. E.J.S. is supported by a Stand Up To Cancer Dream Team award by the Prostate Cancer Foundation, Movember, and Stand Up To Cancer, Grant No. SU2C-AACR-DT0812. We would like to acknowledge the UCSF Center for Advanced Technology and Laboratory for Cell Analysis Core facilities for sequencing and fluorescence-activated cell sorting efforts, respectively.

### Author contributions

R.D., F.Y.F. and L.A.G. led the conception, design, and interpretation of the project as well as writing of the manuscript with the assistance of all of the coauthors. R.D. and L.A.G. led the design and conducting of CRISPR experiments. R.D., E.A.E., L.N.C., J.C., D.K.D., J.T.S., A.G., J.T.H., S.K. led the design, conducting, and analysis of experiments. R.D., F.Y.F., L.A.G., E.J.S., A.A., M.S., R.S., C.Y., W..S.C., D.A.Q. led the design and analysis of genomics, functional genomics experiments, and clinical data.

### Competing interests

F.Y.F. is a co-founder of PFS Genomics, serves on the Scientific Advisory Board of SerImmune, and has consulted for Astellas, Bayer, Blue Earth Diagnostics, Clovis, Genentech, Janssen, Myovant, Roivant, and Sanofi. A.A. is a co-founder of Tango Therapeutics, Azkarra Therapeutics, Ovibio Corporation; a consultant for SPARC, Bluestar, TopoRx, ProLynx, Earli, Cura; a member of the SAB of Genentech and GLAdiator; receives grant/research support from SPARC and AstraZeneca; A.A. also holds patents on the use of PARP inhibitors held jointly with AstraZeneca which he has benefitted financially (and may do so in the future) through the ICR Rewards to Inventors Scheme. L.A.G. has filed patents on CRISPR functional genomics and is a co-founder of Chroma Medicine. F.Y.F. and L.A.G. have filed a patent application related to targeting prostate cancer driver genes. The remaining authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-24919-7>.

**Correspondence** and requests for materials should be addressed to F.Y.F. or L.A.G.

**Peer review information** *Nature Communications* thanks Rene Bernards and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

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



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

© The Author(s) 2021