HHS Public Access

Author manuscript

Oncogene. Author manuscript; available in PMC 2024 May 18.

Published in final edited form as:

Oncogene. 2024 April ; 43(17): 1303–1315. doi:10.1038/s41388-024-02995-5.

In Vivo Genome-Wide CRISPR Screening Identifies CITED2 as a Driver of Prostate Cancer Bone Metastasis

Juan M Arriaga[1,#,*], **Kacey Ronaldson-Bouchard**[2], **Florencia Picech**[1], **Francisca Nunes de Almeida**[1], **Stephanie Afari**[1], **Houssein Chhouri**[3], **Gordana Vunjak-Novakovic**[2,4], **Cory Abate-Shen**[5,*]

[1.]Department of Molecular Pharmacology and Therapeutics, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, NY USA 10032.

[2.]Department of Biomedical Engineering, Columbia University, New York, NY, USA 10032

[3.]Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY USA 10029.

[4.]Department of Medicine, Vagelos College of Physicians and Surgeons, Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, NY, USA 10032.

[5.]Departments of Molecular Pharmacology and Therapeutics, Urology, Medicine, Pathology & Cell Biology, and Systems Biology, Herbert Irving Comprehensive Cancer Center, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, NY USA 10032.

Abstract

Most cancer deaths are due to metastatic dissemination to distant organs. Bone is the most frequently affected organ in metastatic prostate cancer and a major cause of prostate cancer deaths. Yet, our partial understanding of the molecular factors that drive bone metastasis has been a limiting factor for developing preventative and therapeutic strategies to improve patient survival and well-being. Although recent studies have uncovered molecular alterations that occur in prostate cancer metastasis, their functional relevance for bone metastasis is not well understood. Using genome wide CRISPR activation and inhibition screens we have identified multiple drivers and suppressors of prostate cancer metastasis. Through functional validation, including innovative organ-on-a-chip invasion platform for studying bone tropism, our study identifies the transcriptional modulator CITED2 as a novel driver of prostate cancer bone metastasis and uncovers multiple new potential molecular targets for bone metastatic disease.

[*]**Corresponding Authors**: Cory Abate-Shen, Juan M Arriaga, cabateshen@columbia.edu, juan.arriaga@mssm.edu.
[#]Current Address: Department of Oncological Sciences, Department of Urology, Icahn School of Medicine at Mount Sinai, New York, NY USA 10029.

AUTHOR CONTRIBUTIONS: JMA and CAS conceived and designed the study. JMA collected, assembled and analyzed data. KRB performed and analyzed organ-on-a-chip studies. FP performed and analyzed select mouse studies with assistance from SA. FNA and HC performed select *in vitro* studies and RNAseq. HC also performed western blot studies. GVN supervised organ-on-a-chip studies and writing of the manuscript. JMA and CAS wrote the manuscript. CAS supervised all studies. All authors approved the manuscript.

COMPETING INTERESTS: The authors declare no conflict of interest.

Keywords

Bone metastasis; prostate cancer; in vivo screens; functional genomics; CITED2

INTRODUCTION

Metastasis is the ultimate cause of death in most solid tumor cancers, including prostate cancer. A better understanding of the molecular mechanisms that drive its development may lead to more effective therapies. In prostate cancer, the major site of metastasis is bone [1, 2]. Yet, bone metastasis has been particularly challenging to study given difficulties in generating model systems in which these metastases develop spontaneously in the whole organism [3, 4].

In the past decade, multiple studies have documented the molecular landscape of prostate cancer primary tumors and metastases [5–9]. Among the many genomic alterations that have been described, relatively few are more frequently observed in metastasis versus primary tumors, and even fewer have been functionally linked to metastasis [10]. Moreover, given that most metastatic patients receive treatment before any biopsy has been obtained for molecular analysis, it remains difficult to discern whether the observed genomic alterations are relevant for promoting metastasis or arise as a consequence of treatment. Given that multiple transcriptomic and epigenetic differences have been observed between metastases and primary tumors, their functional modeling can be relevant to identify drivers of metastasis.

New technologies have recently emerged that allow transcriptional modulation of gene expression from the endogenous loci of genes, bypassing potential limitations imposed by the expression of exogenous genes. Indeed, the CRISPR/Cas9 system with a catalytically dead Cas9 enzyme (dCas9) coupled to transcriptional co-activators or co-repressors allows specific sgRNAs to be targeted to a particular promoter of interest, to achieve gene activation (CRISPRa) or inhibition (CRISPRi), respectively [11, 12]. The generation of genome-scale sgRNA libraries has further allowed multiple unbiased screening approaches to study the molecular drivers of complex phenotypes, including metastasis [13]. Although potentially very powerful, such screening approaches have had limited applicability on the genome scale *in vivo,* given the large numbers of tumor cells that need to be infected to maintain their adequate representation in libraries, as well as the complex subclonal dynamics that occur *in vivo* that may lead to significant variability of the relative distribution of sgRNAs [13–17].

In this study we have taken advantage of the relatively large size of the mouse prostate that allows for implantation of large numbers of cells and engineered human non-metastatic prostate cancer cell lines for activation or inhibition of gene expression to perform unbiased epigenetic genome-wide, forward *in vivo* screens of treatment-naïve tumors to identify the drivers and suppressors of metastasis. Our results indicate CITED2 as a novel driver of bone metastasis in prostate cancer.

RESULTS

Generation of CRISPRa and CRISPRi non-metastatic human cell lines for infection with genome-wide libraries.

In order to identify drivers of metastasis, we selected 22Rv1, which was originally obtained from a primary prostatic carcinoma rather than metastatic tissue, as is the case for most other prostate cancer cell lines. 22Rv1 is an AR-positive, castration-resistant human prostate carcinoma cell line derived from the CWR22 xenograft [18, 19], that is inherently non-metastatic and has not been reported to metastasize to distant organs when implanted orthotopically into immunodeficient nude mice [20]. To account for potential subclone variability, we generated two clonal sublines from 22Rv1 cells expressing the necessary components for CRISPR-activation (CRISPRa) or CRISPR-inhibition (CRISPRi) strategies, as previously described [11]. Briefly, the nuclease-dead Cas9 enzyme was stably expressed along with a transcriptional activator complex (known as the sunCas9 system) for CRISPRa, or a transcriptional repressor domain (dCas9-KRAB) for CRISPRi. We confirmed the potential for transcriptional activation or inhibition of these cell lines by transducing lentiviral constructs expressing previously validated sgRNAs targeting *CXCR4* or *GSK3B*, respectively (Figure 1A).

To confirm the inherently non-metastatic nature of the clonal CRISPRa/i derivatives, we labeled these cells with a GFP-Luciferase lentiviral construct, infected them with a non-targeting control sgRNA (sgControl) and implanted them into the prostates of *nude* mouse hosts. At the point of euthanasia (see Methods), mice were dissected and organs were harvested and assayed for the presence of metastatic human cells using three independent approaches.

We used *ex-vivo* fluorescence microscopy to visualize metastasis of GFP-expressing cells in all organs and consistently did not observe any distant lung or bone metastases in any of the mice injected with any of our cell line derivatives (Figure S1A–C, n=17 for CRa Clone1; n=24 for CRa Clone2; n=6 for CRi Clone1 and n=6 for CRi Clone2). Local lymph-node metastases were observed with varying frequencies in each of our cell line derivatives, whereas metastasis to liver and pancreas were observed with low frequency in some cell lines (Figure S1A).

We then employed *ex-vivo* bioluminescence imaging of multiple organs upon terminal dissection of mice and confirmed that luciferase activity was largely confined to the prostates of mice, with low levels sporadically detected at local lumbar lymph nodes but not in lungs or bones (Figure S1B).

Finally, we used a sensitive qRT-PCR assay to amplify human Alu repeats [21] that also confirmed the absence of human DNA at distant organs (Figure S1C), providing further support that none of our cell line derivatives have the inherent ability to give rise to distant metastasis.

Having confirmed that the sgControl infecfed CRISPRa/i prostate cancer lines do not metastasize to distant sites, we sought to use these lines to identify genes that would

promote the development of metastasis when implanted into the prostates of host mice (Figure 1B). We used genome-wide libraries that would allow unbiased identification of genes, rather than limiting our approach to libraries representing biological processes that have been previously linked to metastasis. Specifically, we used genome-wide CRISPRa libraries to identify genes whose overexpression would result in distant metastasis (therefore defined as *metastasis driver genes*), as well as genome-wide CRISPRi libraries to identify genes whose repression would result in metastasis (therefore defined as *metastasis suppressor gene*s).

These CRISPRa and CRISPRi genome-wide v2 libraries [22] each contain 5 sgRNAs targeting each of 18,915 genes (104,450 total sgRNAs) and 1895 non-targeting control sgRNAs. The libraries were packaged into lentivirus and infected into target cell lines at a low multiplicity of infection (0.3 – 0.5) such that the majority of the infected cells receive no more than one sgRNA. CRISPRa and CRISPRi screens were each performed in triplicate in each of the two independent 22Rv1-derived clones, totaling 12 replicate screens, n=6 for CRISPRa and n=6 for CRISPRi.

Library-infected cell lines were selected for 72 hours in puromycin to remove uninfected cells and 10 million cells per mouse were implanted into the prostates of 6–10 *nude* mice per experiment, in three independent experiments. This resulted in 700–840X library coverage, a desirable coverage for *in vivo* screens [11, 13]. Tumors were allowed to grow until mice developed signs of illness (see Methods), at which point they were euthanized, metastases were identified by *ex-vivo* fluorescence microscopy and primary and metastatic tumors were harvested. We paid particular attention to bones, carefully removing muscle and connective tissue surrounding the skeleton and fluorescent regions (as in Figure 1D) were macrodissected.

The primary tumor weights were not different between sgControl- or sgLibrary-infected cells in any of the screens performed (Figure S1D, two-tailed t-test), indicating that infection of the different libraries had no overall impact on tumor growth. Kaplan-Meier curves showed marginal differences in survival, with only CRISPRa libraries in Clone2 and CRISPRi libraries in Clone1 cells having significantly decreased overall survival of mice compared to sgControl infected cells (Figure S1F, $P< 0.05$ log-rank test).

The most notable observation in mice injected with either the CRISPRa or CRISPRi infected cells was the presence of bone or lung metastases, with variable frequencies (Figure 1C,D, Figure S1E). In particular, CRISPRa library infection of CRa-Clone1 cells resulted in bone metastasis in 8% of mice (n= 2/24) and in lung metastasis in 17% of mice (n= 4/24), whereas injection of CRa-Clone2 cells resulted in lung metastasis in 18% of mice (n= 5/28). Infection of CRISPRi libraries in CRi-Clone1 cells resulted in bone metastasis in 9% of mice (n=2/23) and lung metastasis in 26% of mice (n=6/23), whereas infection of CRi-Clone2 cells resulted in bone metastasis in 5% of mice (n=1/21). These results show that infection of CRISPRa or CRISPRi libraries into tumor cells that are inherently non-metastatic leads to the development bone or lung metastasis.

Identification of enriched sgRNAs in CRISPRa and CRISPRi metastasis screens.

Next, we sought to identify the metastasis driver and suppressor genes from the screens described above by next-generation sequencing of the sgRNAs present in bone and lung metastases. To this end, we extracted genomic DNA from macrodissected bone and lung metastases, and from whole primary tumors, and amplified the sgRNAs from each sample using a nested PCR approach that added sample barcodes and sequencing adaptors to allow for pooled sequencing.

Of all the sgRNAs present in our libraries, 38–80% were still detectable in primary tumors at the end of each replicate experiment, suggesting a high but variable conservation of sgRNAs during tumor growth and consistent with previous studies showing loss of sgRNA representation as tumors grow *in vivo* [13]. In contrast, we detected 0.1–5.1% of sgRNAs in metastases (Figure S2A), showing high rates of selection compared to primary tumors. We confirmed that over 99% of all the sgRNAs were present in both the initial plasmid library used to generate lentivirus as well as in the infected cell lines prior to orthotopic implantation ("t0 cells") (Figure S2B), with the distribution of sgRNAs being highly correlated between both samples (Pearson $r= 0.78$, $P<0.001$). The correlation of read distributions between replicate experiments was low both in primary tumors and in metastasis, as shown for the CRa Clone1 experiment in Figure S2C, suggesting high variability of subclonal growth dynamics.

Whereas most of the sgRNAs in primary tumors were homogeneously represented at low levels (<5% of total reads per sample) (Figure 2B, Figure S2B), in metastases we identified 1–6 sgRNAs that represented more than 5% of all reads in a particular metastatic sample ("top-represented" sgRNAs) (Figure 2B, C–F), suggesting that metastasis was driven by only a few genes in each experiment. Indeed, none of the 1895 non-targeting sgRNA controls that were present in each library were detected at these levels in metastatic samples, suggesting that the identification of sgRNAs in metastases represent true 'hits' and are not due to chance. The identity of these hits is shown in Figure 2C–F along with their representation in each replicate experiment.

We next compared the distribution of each sgRNA in bone or lung metastases to primary tumors using the Mageck-RRA computational analysis tool [23]. This analysis identified hundreds of significantly enriched sgRNAs (FDR $p<0.001$) (Table S3, Figure S2D) for every condition. Importantly, 28 of the 32 top-represented sgRNAs in metastases were also enriched in metastases compared to primary tumors (Figure S2D), implying that their selection was not due to overrepresentation in the primary tumor. The exception to this was noted in CRa Clone2 lung metastases, where four out of the six top-represented sgRNAs in lung metastasis (ie, targeting *ARF3*, *B3GAT3*, *FBXO33* and *SCRT1*) were not differentially enriched compared to metastases.

While we observed little overlap between the identified sgRNAs in different experiments (Figure 2G), in CRa Clone1 31% of significant sgRNAs in bone metastasis were also significant in lung metastasis (Figure 2G) and one of these (targeting *CITED2*) was also significantly enriched in lung metastases from Clone 2. Moreover, gene-level analysis considering all 5 sgRNAs targeting every gene also identified *CITED2* as a significant hit,

with 3 of the 5 sgRNAs significantly enriched in the CRa Clone1 Bone metastasis screen (Figure 2H, FDR-adjusted p=0.0025). We thus decided to focus our validation studies on CITED2.

CITED2 drives bone metastasis in prostate cancer.

To validate CITED2 as a potential driver bone metastasis in our CRISPRa screens, we analyzed the consequences of CITED2 expression in bone metastasis using *ex vivo* and *in vivo* approaches. First, we used novel organ-on-a-chip assays to study invasion of cell lines into bone [24]. In these assays, primary tumor cells are introduced into a chip where a circulating flow is separated from a chamber containing engineered bone by engineered vascular barriers. Cells are cultured in this system, allowed to migrate through the endothelium lining into the bone chamber, and the resulting invading cells are quantified by their fluorescence emission (Figure 3A, see Methods).

In addition to the 22Rv1-CRa Clone 1 cells used for our screen, we overexpressed CITED2 cDNA in a second human prostate cancer cell line, LNCaP. Further, given the scarcity of human models derived from primary tumors, we also used mouse cell lines derived from non-metastatic primary tumors of the NP (*Nkx3–1*^{CreERT2}; *Pten*$^{fl/fl}$) and NPM (NP-HiMyc) genetically engineered mouse models of prostate cancer (Figure S3A,B) [25, 26]. After 4 weeks of culture in this system, invasion of cells into the bone compartment was significantly higher in all CITED2-overexpressing cells compared to their respective controls (Figure 3A, $p<0.05$ two-tailed t test), indicating that CITED2 has a functional role in promoting invasion to bone. Representative histological sections of the invaded engineered bones from 22Rv1-CRa Clone 1 cells are shown in Figure 3A.

We further tested whether CITED2 would favor growth of tumor cells in bone by intratibial implantation of 22Rv1-CRa Clone1 CITED2 cells into SCID mice. Indeed, tumors with CITED2 overexpression grew significantly larger than controls after 6 weeks growth as shown in Figure 3B (p=0.0079, two-tailed Mann Whitney test, n=5). Representative histological sections of the resulting tumors are shown in Figure S3C.

We confirmed that individual expression of CITED2 into non-metastatic 22Rv1 cells is sufficient for development of bone metastasis by orthotopic implantation of 22Rv1-CRa Clone1 CITED2 cells into nude mice. Indeed, three out of eight injected mice showed histologically confirmed bone metastasis in the spine compared to none of 12 sgControl injected tumors (Figure 3C, p=0.049, two-tailed Fisher's Exact test), despite no difference in primary tumor growth (Figure S3D). Representative sections of the resulting tumors are shown in Figure 3C and Figure S3D.

Lastly, we explored whether targeting CITED2 could be a feasible strategy to inhibit bone metastasis by knocking down CITED2 using CRISPRi in cell lines derived from our highly bone metastatic, syngeneic NPK mouse model (Figure S3A,B). We then injected sgControl and sgCITED2 infected cells intracardially into immunodeficient nude and immunocompetent C57/Bl6 mice and identified bone metastasis by *ex vivo* epifluorescence imaging two weeks after injection. Indeed, knockdown of CITED2 resulted in significant

inhibition of bone metastasis in both nude and C57 mice, as shown in Figure 3D, suggesting that CITED2 may be both necessary and sufficient for bone metastasis in prostate cancer.

CITED2-driven transcriptional profiles differ in primary and metastatic prostate cancer.

We then studied whether expression of CITED2 was associated with metastasis in human prostate cancer tissues. We first confirmed that CITED2 is expressed in clinical samples of prostate cancer bone metastasis by immunohistochemical staining. Notably, as shown in Figure S4A, CITED2 was ubiquitously expressed in prostate cancer epithelium of both bone metastasis (n=4) and primary tumors (n=10, radical prostatectomies). CITED2 mRNA expression in publicly available datasets [9] was also not different in metastases and primary tumors, although higher levels of expression can be observed in a subset of both soft tissue and bone metastases (Figure S4B). Given that many factors may confound gene expression analysis in human tissues, including prior exposure to different treatment modalities, we also analyzed CITED2 expression in the NPK genetically engineered mouse model of spontaneous, bone metastatic, treatment-naive mCRPC [26]. Remarkably, in this context *Cited2* is expressed at higher levels in bone metastasis than in primary tumors or other metastatic sites (Figure S4C, $p<0.0001$, one-way ANOVA with Dunnett's multiple comparisons test, compared to prostate). We then assessed whether mRNA levels in tumors were associated with patient survival. As had been shown before [27], CITED2 expression was significantly associated with worse overall survival in the cancer genome atlas cohort (TCGA [5]) of primary tumors (Log-rank test P=0.0245, Figure S4D), although not with disease-free survival or with overall survival in metastatic patients from the stand up to cancer (SU2C) cohort [9](Figure S4D). Overall, our data implies that CITED2 expression may be highly expressed in a subset of metastases and potentially associated with survival in a subset of primary tumors.

CITED2 is a transcriptional co-regulator that does not bind DNA directly but is able to modulate transcriptional programs [28–31]. Therefore, we hypothesized that modulation of its transcriptional 'activity' (rather than overall expression levels) may be associated to development of bone metastasis in prostate cancer. To assess whether CITED2-driven transcriptional profiles were altered in bone metastasis compared to primary prostate tumors, we inferred CITED2 activity by performing genome-wide correlation of mRNA expression to CITED2 mRNA levels in metastatic (SU2C, FHCRC[7]) or primary tumor (TCGA, DFKZ[32]) human prostate cancer patient datasets. We ranked all genes according to Spearman correlation coefficients and performed preranked GSEA analysis using the Hallmarks gene sets, to thus identify pathways significantly associated with CITED2 expression levels in these two clinical scenarios (Figure 4A). Interestingly, the pathways associated with CITED2 expression levels in metastatic datasets were similar to each other but markedly different, even opposite, from primary tumors datasets, indeed suggesting that CITED2 activity may be different between metastases and primary tumors.

To study transcriptional programs regulated by CITED2 in human prostate cancer cells, we next performed RNAseq on 22Rv1 CRa Clone1 cells infected with sgCITED2 and sgControl. We also included cDNA overexpression on LNCaP cells, another human cell line that does not develop distant metastasis when injected in nude mice [33] (Figure 4B–

D, S4E–H). Our analysis identified 38 upregulated and 39 downregulated genes in 22Rv1 cells and 165 upregulated and 392 downregulated genes in LNCaP cells, with minimal overlap in both cell lines despite similar levels of CITED2 overexpression (LFC>0.3 and Padj<0.01, Figure S4E). Individual expression of differentially expressed genes in 22Rv1 cells is shown for metastatic datasets (SU2C and FHCRC) in Figure S4F. We then compared these gene signatures in both cell lines to the top 200 differentially expressed genes in a patient-derived gene signature comparing bone metastases (SU2C, n=74) to primary tumors (TCGA, n=497) using GSEA. Remarkably, upregulated genes in metastases were significantly positively enriched with CITED2-regulated genes in 22Rv1 cells (NES 1.99, FDR<10^{-5}), but significantly negatively enriched with those of LNCaP cells (NES −1.7, FDR<5×10^{-4}). Downregulated genes in metastases were significantly positively enriched with CITED2-regulated genes in LNCaP cells (NES 1.65, FDR<10^{-5}) and not significantly changed in 22Rv1.

We next performed GSEA using the c2-c8 (Figure 4C, S4G) and Hallmarks (Figure 4D) gene sets from the MSigDB database to identify the signaling pathways associated to CITED2 overexpression in 22Rv1 and LNCaP cells (FDR <0.1). Notably, Hallmarks pathways differentially expressed in metastasis were consistent with pathways altered in 22Rv1 cells (Figure 4D, first two columns) and opposite to those altered in LNCaP cells (Figure 4D, columns one and three), with E2F targets and G2/M checkpoint standing out as top pathways. Moreover, most gene sets containing the term “metastasis” in the c2-c8 MSigDB pathway database were enriched in 22Rv1-CITED2 cells, whereas the opposite occurred in LNCaP-CITED2 cells (Figure 4C). Given that our organ-on-a-chip studies (Figure 3A) had showed that CITED2 overexpression promoted invasion to bone in LNCaP cells, we hypothesized that CITED2 could impact later stages of metastasis (ie, dormancy or colonization) that could explain the inverse correlation to metastasis signatures. We thus searched the c2-c8 database for pathways containing the term ‘quiescence’ and indeed observed that CITED2 overexpression in LNCaP cells favored quiescence signatures, which were conversely repressed in 22Rv1 cells (Figure S4G). We further included a dormancy signature derived by Kim et al [34] to confirm the opposite effects of CITED2 in these cell lines, and show that genes associated with dormancy were inversed by CITED2 overexpression in 22Rv1 cells (both upregulated and downregulated genes in the signature), whereas in LNCaP cells CITED2 significantly repressed downregulated genes of the dormancy signature (Figure S4G). Furthermore, although LNCaP-CITED2 cells grew steadily in culture, colony formation assays of cells plated at low density confirmed that CITED2 overexpression repressed cell proliferation in LNCaP cells, and conversely favored proliferation in 22Rv1 cells (Figure 4E and Figure S4H). We also tested whether CITED2 could alter proliferation upon standard of care enzalutamide treatment, although we did not observe any changes in both treatment-resistant 22Rv1 cells and treatment-sensitive LNCaP cells (Figure S4H).

Overall, our results show that CITED2 overexpression has distinct transcriptional effects on different cells lines, with effects on 22Rv1 resembling those observed in metastatic datasets and those on LNCaP cells resembling primary tumors.

CITED2 and E2F1 co-expression identify a subset of prostate tumors with worse survival.

Among the top pathways enriched in 22Rv1-CITED2 cells (Figure 4D) and in metastasis versus primary tumors (Figure 4D) was the E2F targets gene set, a known regulator of cell proliferation. We thus analyzed which E2F1 family members could be associated with CITED2 regulation of metastasis by studying their differential expression on 22Rv1 and LNCaP cells, and upon CITED2 overexpression. We found that *E2F1* was not only significantly lower in LNCaP cells compared to 22Rv1 cells, but that CITED2 overexpression repressed *E2F1* mRNA expression only in the former (Figure S4H). Western blot analysis was consistent with these results, showing that CITED2 overexpression could repress E2F1 in LNCaPs and conversely stimulate E2F1 in 22Rv1 cells (Figure 4F). Thus, since CITED2 overexpression has opposite effects on metastasis-associated signatures and on E2F1 expression in 22Rv1 and LNCaP cells, we hypothesized that E2F1 expression could be used to distinguish survival outcomes in patient datasets. We thus analyzed patient survival in publicly available datasets by stratifying cases into CITED2-low and -high expression as well as into E2F1-low and -high. Our results show that high E2F1 expression was significantly associated with worse disease-free survival in TCGA primary tumors and with overall survival in SU2C metastases only in CITED2-high, but not in CITED2-low cases (Figure 4G). Thus, these results are consistent with E2F1 as a mediator and/or biomarker of the pro-metastatic effects of CITED2, which are observed upon overexpression in 22Rv1, but not in LNCaP, cells.

DISCUSSION

Whereas localized prostate cancer has a favorable prognosis, metastatic disease is invariably lethal [35]. The main site of metastasis from prostate cancer is bone and yet progress in our understanding of bone metastasis has been hampered by the lack of model systems that recapitulate this phenotype [3].

In this study, we have undertaken an unbiased approach to identify drivers and suppressors of prostate cancer metastasis, leveraging *in vivo* screens and genome-wide activation and inhibition of gene expression from endogenous loci. By engineering two clonal derivatives of a non-metastatic human prostate cancer cell line, we observed that expression of both activation and inhibition libraries leads to metastasis to bone or lungs with variable frequencies. Our screens identified several driver sgRNAs and most notably highlight CITED2 as a novel driver of bone metastasis for prostate cancer, which we propose has different transcriptional activity between metastasis and primary prostate tumors that can be distinguished by concomitant high expression of E2F1. Furthermore, we used state-of-the-art bone engineering models as part of our functional validation strategy, showcasing a novel approach to study invasion of tumor cells into humanized bone.

In our screens, the correlation of sgRNA distributions was low in all tumor samples, likely reflecting a highly variable subclonal growth in *in vivo* conditions. This can account for both the variable frequencies of distant metastases as well as for the little overlap of the identity of individual sgRNAs present in metastatic samples, across replicates, clonal derivatives and organ sites. Given that library coverage was high in the initial implantation of tumors as well as (in many cases) after 2–3 months of *in vivo* growth, our results suggest that factors other

than cell number at implantation can influence the metastasis outcomes of *in vivo* screens. Of note, CITED2 overexpression resulted in bone metastasis with incomplete penetrance, highlighting the inefficient nature of metastasis even when cell intrinsic drivers of metastasis are present, as has also been observed in other prostate bone metastasis models reported by us and others [26, 36].

Moreover, cell-extrinsic factors such as interaction with different host cells as well as positional cues such as proximity to blood vessels, lymph vessels, nerves or hypoxic regions might also influence the efficiency of metastasis and therefore screen outcomes [37, 38]. Taken together, these results suggest that genome-wide *in vivo* screens have the potential to identify genes that are biologically relevant to metastasis, although they may not necessarily lead to the identification of all possible metastasis drivers and suppressors.

Our results align with prior reports showing that CITED2 can promote metastasis [27, 39, 40]. We show that CITED2 may act at multiple levels of the metastatic cascade [41] since its overexpression in 22Rv1 cells: i) results in higher invasion to bone in state-of-the-art organ-on-a-chip assays, ii) shows higher proliferation in bone by intratibial assays, and iii) is sufficient for the occurrence of bone metastasis in orthotopic prostate xenografts. Moreover, inhibiting CITED2 in a highly metastatic mouse model can impair bone metastasis, proposing that CITED2 can serve as a promising therapeutic target.

CITED2 is a transcriptional modulator that may regulate many pathways that are crucial for cancer development and progression, including p53 [31], Myc [29], TGF-B [30] and hypoxia [28] signaling pathways. Here we provide evidence that the transcriptional activity of CITED2, but not its overall expression, is different in prostate cancer metastasis than in primary tumors. In tumors like the 22Rv1 cell line, CITED2 overexpression mimics gene and pathway alterations enriched in metastasis versus primary prostate cancer specimens, whereas in tumors like LNCaP cells, CITED2 mimics primary tumors. We further show that activation of the E2F pathway underlies at least in part these differential effects on metastatic propensity, and that combined high CITED2 and high E2F1 expression may identify such tumors associated with poorer patient survival. Indeed, opposite effects of CITED2 on the E2F pathway have been reported in different cell types. CITED2 can stimulate MYC/E2F-regulated proliferation in lung cancer [29], but repress the E2F pathway in hematopoietic stem cells, favoring their quiescence[42]. Thus, identifying the factors that mediate these opposing effects will be important to further understand the role of CITED2 in cancer progression.

In a previous study [27], Shin et al showed that CITED2 may promote metastasis of prostate cancer through activation of the nucleolin-Akt pathway. CITED2 expression was higher in prostate cancer than in the benign prostate, correlated with Gleason grade and was associated with worse survival outcomes. Other studies have showed that CITED2 may mediate bone metastasis of breast cancer, through modulation of the NFKB pathway [39, 40]. Our study complements these findings and, to our knowledge, is the first to report both expression of CITED2 in human specimens of bone metastasis from prostate cancer, and a direct role for CITED2 in mediating prostate cancer metastasis into bone. This is important given that bone metastasis from different cancer types may have different

molecular underpinnings [43]. Identifying shared mechanisms may thus lead to more broadly applicable therapeutic strategies.

In summary, our genome-wide *in vivo* screens have identified a subset of all possible drivers and suppressors of metastasis and established a functional role for CITED2 in enabling prostate cancer metastasis to bone.

MATERIALS AND METHODS

Cell lines and generation of CRISPRa/i derivatives.

22Rv1 (ATCC, Manassas, VA CRL-2505) and LNCaP (ATCC, CRL-1740) cells were grown and maintained in RPMI 1640 (ATCC modification, Thermo-Fisher A1049101) supplemented with 10% FBS (Sigma-Aldrich, St. Louis, MO, F0392), whereas HEK-293FT (Invitrogen Waltham, MA, R700–07) were cultured in DMEM-10%FBS (Thermo Fisher Scientific, Waltham, MA,11–995-073). Cell lines were authenticated by STR profiling, passaged twice-weekly, used within 30 passages and tested negative for Mycoplasma using Universal Mycoplasma Detection Kit (ATCC #30–1012K).

Clonal cell derivatives capable of target gene overexpression by CRIPSRa or repression by CRISPRi were generated as previously described [11]). Briefly, to generate 22Rv1-CRa derivatives, parental cells were first transfected with lentivirus generated using the pHRdSV40-dCas9–10xGCN4_v4-P2A-BFP plasmid (Addgene, Watertown, MA #60903), and sorted for blue-fluorescence using a BD FACS Aria II sorter (BD Biosciences, Franklin Lakes, NJ). Next, sorted cells were infected with lentivirus generated using the pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS plasmid (Addgene #60904), and sorted for blue and green-fluorescence. The resulting cells were plated as isolated clones in 96-well plates and multiple clonal derivatives were expanded, infected with sgCXCR4 (positive control [12]) or sgControl lentivirus and selected with 2ug/ml puromycin (Sigma-Aldrich) for 7 days. After 7 days, clones were dissociated in 10 mM EDTA/PBS and stained in PBS/10% FBS for 1 hr at room temperature with APC labeled anti-CXCR4 antibody (Anti-Mouse CD184 (CXCR4) APC, (eBioScience San Diego, CA)). The clones showing >15-fold expression of endogenous CXCR4 levels in >85% cells were selected for further use.

To generate 22Rv1- and NPK-CRi derivatives, parental cells were transfected with lentivirus generated using the pHR-SFFV-KRAB-dCas9-P2A-mCherry plasmid (Addgene #60954), sorted for mCherry-fluorescence and plated as isolated clones in 96-well plates. Cells were infected with sgGSK3B or sgControl cells and treated as for CRISPRa derivatives and at 14 days after puromycin selection, clones showing >75% inhibition of endogenous *GSK3B* mRNA levels (measured by qRT-PCR) were selected for further use.

CRISPRa/i derivatives were engineered to express luciferase by lentiviral transduction of the pHAGE PGK-GFP-IRES-LUC-W plasmid (Addgene #46793) and sorted for GFP-positive cells by flow cytometry.

Mouse NP cell lines were derived from primary tumors of *Nkx3–1*^{CreERT2}; *Pten*$^{flox/flox}$ mice as described in [25] whereas NPM cells were derived from Nkx3–1^{CreERT2}; *Pten*$^{flox/flox}$;

Pb-HiMyc; R26$^{CAG\text{-}LSL\text{-}EYFP}$ mice [26] and NPK cells from Nkx3–1^{CreERT2}; *Pten*$^{flox/flox}$; *Kras*^{LSLG12D}; R26$^{CAG\text{-}LSL\text{-}EYFP}$ mice [26]. All of these cells were grown in RPMI 1640 supplemented with 10% FBS.

RNA extraction and qRT-PCR analysis

RNA from cell lines was extracted with TRIZOL reagent (Thermo Fisher, 15–596-026), reverse-transcribed with SuperScript™ III First-Strand Synthesis SuperMix (Thermo Fisher 11–752-050), and RNA expression measured by quantitative real time PCR (qRT-PCR) using the QuantiTect SYBR Green PCR kit (Qiagen, Germantown, MD, 2045). Sequences of all primers used in this study are provided in Table S1.

Amplification of sgRNA libraries

Genome-wide libraries (CRISPRa and CRISPRi v2 libraries with top5 sgRNAs/gene, Addgene #83978 and # 83969, respectively) were a kind gift from Dr Jonathan Weissman and were amplified by electroporation of 100ng DNA into Endura electrocompetent cells (Lucigen Corporation, Middleton, WI #60242) using an Amaxa Nucleofector II device (Lonza, Cologne, Germany) and Fisherbrand #FB101 electroporation cuvettes. After shaking at 225rpm for 1 hour at 37°C in recovery media, cells were transferred to 500ml LB medium supplemented with 50 ug/ml carbenicillin (Sigma-Aldrich) and grown for 14 hours at 32°C. Pellets were harvested and plasmid DNA purified using NucleoBond® Xtra Maxi EF (Macherey-Nagel Inc, Allentown, PA #740426). High transformation efficiency (>10^{11} colony forming units/ug) was confirmed using serial dilutions into LB-agar plated supplemented with 50ug/ml ampicillin (Sigma-Aldrich).

Cloning of CRISPRa, CRISPRi and lentiviral overexpression constructs

For individual sgRNA expression, the pU6-sgRNA EF1Alpha-puro-T2A-BFP vector (Addgene #60955) was digested with BstXI and BlpI enzymes (NEB, New England Biolabs, Ipswich, MA) and ligated to 100nM annealed oligonucleotides carrying the sgRNA sequence of interest with Mighty Mix DNA Ligation Kit (Takara Bio, San Jose, CA). For this purpose, sense (5’ ttg[X]gtttaagagc 3’) and antisense (5’ ttagctcttaaac[Y]caacaag 3’) oligonucleotides were synthesized, where [X] corresponds to the target sgRNA sequence and [Y] to its reverse complementary sequence, and 10uM sense and antisense oligonucleotides were annealed and phosphorylated with T4 PNK enzyme (NEB).

Lentiviral expression vectors for CITED2 were generated by Gibson cloning (NEBuilder® HiFi DNA Assembly Cloning Kit, NEB E5520S). Human CITED2 cDNA was cloned into the PLX304 constitutive expression plasmid (Addgene #25890) with an N-terminal V5 tag or into the RFP-shRNA locus of the inducible pTRIPZ vector, with an HA tag. Inducible expression was achieved by using 0.5ug/ul doxycycline (Sigma-Aldrich, D9891).

Lentivirus production and transduction

HEK-293FT cells were plated at ~90% confluency into 150mm dishes 24 hours before transfection, in DMEM-10%FBS culture media. 30ug lentiviral vector (pU6-sgRNA EF1Alpha-puro-T2A-BFP, Addgene #60955 expressing different sgRNA sequences, see Table S1), 24ug psPAX2 (Addgene #12260) and 12ug pMD2.G (Addgene #12259)

packaging plasmids were diluted in OPTIMEM-I media (ThermoFisher) and mixed with 60ug/ml polyethilenimine (Sigma-Aldrich) for transfection into HEK-293FT cells. After overnight incubation, the media was replaced with DMEM-1%FBS for 48hours, centrifuged at 300g for 5 minutes and filtered through a 0.45um filter to obtain the virus supernatant that was mixed with 0.8 ug/ml polybrene (Sigma-Aldrich) to infect target cells. For pooled CRISPRa/i library transfection, cells were infected at 30–40% infection rate in order to achieve a multiplicity of infection of 0.3 – 0.5, such that most cells would receive no more than one sgRNA. For antibiotic selection, the media was changed 48 hours after viral infection to RPMI-10%FBS supplemented with 2ug/ml puromycin and selected for 72 hours (for genome-wide infected cells) or for 7 days for individual plasmid infection. Individual CRISPRa and CRISPRi sgRNAs were evaluated for target gene activation or repression, respectively, at >10 days after virus infection.

Mouse studies

All experiments using animals were performed according to protocols approved by the Institutional Animal Care and Use Committee (IACUC) at Columbia University Irving Medical Center. 6–8-week-old male athymic nude-Foxn1^{nu} mice were purchased from Envigo (Boyertown, PA). Since the focus of the study is prostate cancer, only male mice were used. All mice were housed in pathogen-free barrier conditions under 12-h light–dark cycles and with temperature and humidity set points at 20–25 °C and 30–70%, respectively. For orthotopic implantation mice were anaesthetized by intraperitoneal injection of ketamine (100mg/kg) and xylazine (16mg/kg) (Patterson Vet Supply, Greely, CO) and an incision was made in the lower abdomen under IACUC approved protocols.

For the genome-wide screen experiments, tumor cells were resuspended in 50% Matrigel (Thermo Fisher) in PBS and 10 million cells were injected into the anterior prostates, using 50ul and a 30-gauge ½ inch needle on a Hamilton syringe. For validation experiments, 1 million cells were used. The incision site in the muscle layer was then closed with surgical suture (Ethicon J385H Vicryl absorbable suture, size 5–0) and the skin incision by wound clips.

Tumors were allowed to grow for 2 to 4 months, until their body condition score [44] was <1.5, showed signs of genitourinary occlusion or signs of distress. Tumor growth was monitored *in vivo* and metastatic dissemination *ex vivo* by bioluminescence imaging using an IVIS Spectrum Optical Imaging System (Perkin Elmer, Waltham, MA). Ten minutes prior to imaging, mice were injected intraperitoneally with 150mg/kg D-luciferin (Perkin Elmer) for *in vivo* analysis, whereas for *ex vivo* analysis luciferin was used at 300ug/ml. Images were generated and quantified using Living Image Software (Perkin Elmer). At the time of sacrifice, GFP-positive prostatic tumors and metastases were visualized by *ex vivo* fluorescence using an Olympus SZX16 microscope (Ex490–500/Em510–560 filter).

Genomic DNA extraction, library preparation and targeted DNA sequencing

For isolation of genomic DNA, freshly dissected metastases and primary tumors from tumor-bearing mice were macrodissected under *ex vivo* fluorescence microscopy, weighed and genomic DNA then isolated using Zymo Quick gDNA MidiPrep kit (Thermo Fisher).

Pooled sgRNA libraries were amplified with a nested PCR strategy using high-fidelity Phusion Flash polymerase (Thermo). PCR#1 was performed using “Library Amplification PCR1” primers (Table S1) using 25 cycles and annealing at 55°C for 5 seconds. Input DNA was 1ug for samples obtained from metastases and 200ng for samples from primary tumors or plasmid DNA. PCR#2 was used to add sample-specific barcodes as well as to prepare the libraries for next-generation sequencing, using “Library Amplification PCR2” primers from Table S1, 18 amplification cycles, annealing at 56°C for 5 seconds and $1/10^{th}$ volume of PCR#1 as input.

In order to maintain 1000X coverage of every sgRNA (ie, n= 1×10^8 sgRNAs/library) at the PCR amplification steps, assuming 6.6pg of DNA/cell [13], we estimated to need 660ug of genomic DNA (gDNA) as input for every sample (ie, 660 PCRs using 1ug as input). For primary tumors, gDNA extractions contained mostly tumor-derived DNA and therefore, for every replicate experiment, gDNA was pooled from all primary tumors proportionally to their individual tumor weights and 660ug were used directly as input for PCR#1. For metastases, gDNA extraction would contain variable amounts of DNA from the host mouse tissue and therefore the total amount of human-derived gDNA in a given sample was estimated based on qPCR quantification of sgRNA abundance, using Library Amplification PCR1 primers. For this purpose, a calibration curve was generated by spiking-in variable amounts of sgControl-infected 22Rv1-CRa cells in a total of 1ug gDNA obtained from lung tissue of a non-injected *nude* mouse. This yielded a linear detection range (R^2=0.9525) from 1–200ng input DNA. Lung and bone metastasis derived gDNA was independently pooled from all corresponding metastases in a given replicate experiment, and 660ug of estimated human-derived gDNA were then amplified by PCR.

All final PCR2 products were visualized on a 2% agarose gel using SYBR-Safe stain (Thermo Fisher) and DNA from positive samples was purified with SPRI beads (Beckman Coulter Inc., Indianapolis, IN), quantified using a Qubit 4 fluorometer (Thermo Fisher) and denatured in NaOH 0.1N before loading into an Illumina® NextSeq® 500 system at 1.08pM with 20% PhiX spike-in.

After sequencing, we used MAGeCK-RRA[23] (https://sourceforge.net/p/mageck/wiki/Home/) to generate count tables (Table S2) and to perform differential expression of sgRNA abundance using total read count normalization and pairwise comparisons between metastasis and primary tumor replicates. This yielded both sgRNA- and gene-level differential expression shown in Tables S3 and S4, respectively. Downstream statistical analysis and figures were generated using normalized read counts in R v.3.6.3 and R-studio 1.3.959–1 or GraphPad Prism v9.

Histological Analysis and Immunohistochemistry

For histological analyses, tissues were fixed in 10% formalin (Fisher Scientific, Fair Lawn, NJ) for 48hs. Bones were then decalcified for three weeks in 15% EDTA pH7.0 solution (Sigma-Aldrich) and changed daily. Histopathological analyses were done using 3 μm paraffin sections and hematoxylin and eosin (H&E) staining. Immunohistochemistry was performed with citrate-based antigen retrieval (Vector Labs, Newark, CA, H-330) using the VECTASTAIN® Elite® ABC HRP Kit (Vector Labs, PK-6100). Antibodies used were

anti-CITED2 mouse monoclonal (1/50, Invitrogen MA316523) and horse anti-mouse IgG (1/10.000, Vector Labs BA-2000).

Unpublished cohorts used anonymized human tissue specimens from Columbia University Irving Medical Center (CUIMC), including radical prostatectomy samples from primary tumors and 4 bone metastasis biopsies from patients with prostate cancer. All studies using human tissue specimens were performed according to protocols approved by the Human Research Protection Office and Institutional Review Board at CUIMC.

RNAsequencing

RNA was prepared from snap-frozen cell line pellets using the MagMAX-96 total RNA isolation kit (Thermo Fisher Scientific). Total RNA was enriched for mRNA using polyA pull-down; only RNA samples having between >200 ng and 1 μg and with an RNA integrity number >9 were used. Libraries were made using an Illumina TruSeq RNA chemistry and sequenced using an Element AVITI at Columbia Genome Center at >20 million paired-end 75bp reads for each sample, in duplicate. We used bases2fastq version 1.1.0.593880262 for converting BCL to fastq format, coupled with adaptor trimming. We performed a pseudoalignment to a kallisto index created from transcriptomes (Ensembl v96, Human:GRCh38.p12) using kallisto (0.44.0). We tested for differentially expressed genes using DESeq2.

Western Blotting

Total protein extracts were prepared by lysis of cells with 1X radioimmunoprecipitation assay (RIPA) buffer (0.1% SDS, 1.0% deoxycholate sodium salt, 1.0% Triton X-100, 0.15 M NaCl, 10 μmol/L Tris-HCl (pH 7.5), 1 mmol/L EDTA) with fresh 1% protease inhibitor (#1697498; Roche Basel), PMSF (10837091001, Sigma-Aldrich) and 1% phosphatase inhibitor (#P2850; Sigma-Aldrich). Protein lysates (20 μg per lane) were resolved by SDS page, followed by immunoblotting with primary antibodies (anti-CITED2 1/500, Invitrogen MA316523, anti-V5 1/500, Thermo-Fisher R96025 and anti-B-actin 1/20000 Cell Signaling cs4970, anti-E2F1 1/1000, Life Technologies, Grand Island NY 66515–1-Ig), secondary antibodies (anti-Rabbit IgG HRP Sigma NA934 or anti-mouse IgG HRP Sigma NA931, 1/10000) and visualized using an ECL Plus Western Blotting Detection Kit (Thermo-Fisher RPN2232).

Organ-on-a-chip Invasion Assay

Engineered bone tissues were made from bovine calf metacarpals (Lampire Biological Laboratories, 19D24003) that were sectioned into rectangular scaffolds (4 mm wide x 8 mm long x 1 mm thick) and fully decellularized as detailed previously [24, 45] The decellularization process removed all bovine cellular components, leaving just the bone extracellular matrix and bone architecture. The bones were lyophilized and the only bone scaffolds weighing between 12 and 18 mg/scaffold were included, resulting in the starting material with uniform porosity for engineering bone tissue. After sterilization in 70% ethanol overnight and 24 hours of incubation in Dulbecco's Modified Eagle Medium (DMEM) the bone scaffolds were seeded with human cells.

Human bone-marrow derived mesenchymal stem cells (MSCs) (Lonza) were infused into the bone scaffolds (4×10^5 cells per scaffold) by suspending the cells in 40 μL of medium (DMEM supplemented with 10% (v/v) HyClone fetal bovine serum (FBS), 1% penicillin/streptomycin, and 1 ng/mL of basic fibroblast growth factor, bFGF), and letting the cells attach for 2 hours before adding additional media (2 mL per well). To support the differentiation of the MSCs into osteoblasts within the engineered bone matrix, the MSC seeded bone was cultured in osteogenic medium consisting of low glucose DMEM supplemented with 1 μM dexamethasone (Sigma-Aldrich), 10 mM β-glycerophosphate (Sigma-Aldrich), and 50 μM L-ascorbic acid-2-phosphate (Sigma-Aldrich).

The osteogenic differentiation process continued for three weeks, with media changes three times a week. To create osteolytic bone, we then infused primary monocytes into the osteoblastic bone scaffolds [45]. CD14+ monocytes were obtained by negative selection (EasySep Human Monocyte Isolation Kit, Stem Cell Technologies, 19359) from peripheral blood mononuclear cells (PBMC) isolated from buffy coats of human blood (fully deidentified samples obtained from the New York Blood Center) via density gradient centrifugation with Ficoll-Paque PLUS (GE Healthcare, 17–1440-02). Purified monocytes were seeded into the engineered bone tissues (4×10^5 cells per scaffold) by suspending the cells in 40 μL of osteolytic medium for two hours. The engineered bone tissues were then cultured for a week in 2 mL of osteolytic media (Minimum Essential Medium Eagle Alpha modification (α-MEM, Sigma, M4526) supplemented with 10% (v/v) heat-inactivated HyClone FBS, 1% penicillin/streptomycin, l-Glutamine (Gibco, 25030–081), 20 ng mL^{-1} Recombinant Human M-CSF (PeproTech, 300–25) and 40 ng mL^{-1} Recombinant Human sRANK Ligand (PeproTech, 310–01)), with media changes and fresh cytokines every three days.

The engineered bone tissues were then placed into a chip designed for inter-organ communication that we recently developed [24]. Briefly, engineered vascular barriers are formed by seeding 1.5×10^5 MSC cells and 4×10^5 human umbilical venous endothelial cells (HUVEC) on custom made transwell inserts. After the cells attached to the transwell barrier and formed a confluent monolayer, they were placed into the 1-tissue multi-organ chip and exposed to increasing levels of shear stress (0.31 dynes cm^{-2} for 12 hours, 0.63 dynes cm^{-2} for 24 hours, 1.88 dynes cm^{-2} for 24 hours). The ramping of shear creates a confluent and quiescent vascular barrier lining the bottom of the chamber with engineered bone. After the engineered bone is added to the chamber (directly above the vascular barrier), circulating primary tumor cells were introduced into the vascular reservoir and allowed to circulate underneath the vascular barrier at a hydrodynamic shear stress of 1.88 dynes cm^{-2}).

The cancer cells were labelled (LuminiCell Tracker 670, Sigma-Aldrich) to enable downstream tracking and allowed to circulate for 4 weeks. Media changes occurred every other day by replacing 1 mL of medium from the vascular reservoir with fresh vascular medium (EGM-2, Lonza) and 1 mL of medium from the engineered bone tissue reservoir with osteolytic medium specified above.

Intravasation of circulating cancer cells within the bone tissues was tracked using the IVIS Spectrum Optical Imaging System (Perkin-Elmer) in Columbia University's Oncology Precision Therapeutics and Imaging Core. The chips with engineered tissues were placed next to one another in the same field of view and imaged using an IVIS 200 Spectrum device. The corresponding IVIS Spectrum software (Perkin-Elmer) was used to analyze the images by converting the signal to the normalized Radiant Efficiency (Emission light [photons/sec.cm^2 str]/Excitation light [μW/cm^2]). We measured the fluorescence of the labelled cancer cells within the engineered bone tissues by selecting the same region of interest for each tissue and quantifying the sum of the radiant efficiency of all fluorescent pixels within the region of interest.

Intratibial and intracardiac implantation studies

For monitoring tumor growth in bone, 22Rv1-CRa sgCITED2 and sgControl cells (1 × 10^6 cells in 10 μl of PBS) were injected into the tibiae of male NOD-SCID mice (NOD.CB17-Prkdcscid/J, strain 001303, Jackson Laboratories). A small longitudinal skin incision was made across the knee capsule, the tip of a scalpel was used to drill a hole into which cells were injected, sterile surgical bone wax (CP Medical, Norcross, GA) was used to seal the hole and the skin was then closed with wound clips. Tumor growth was monitored biweekly by bioluminescence imaging using an IVIS Spectrum Optical Imaging System (PerkinElmer), following intraperitoneal injection of 150 mg /kg d-luciferin (PerkinElmer). Images were quantified using Living Image Software (PerkinElmer). For intracardiac metastasis assays, mouse NPK bone metastatic cells (30) (1 × 105 cells in 100 μl of PBS) were injected percutaneously into the left heart ventricle of immunodeficient NCr nude mice (male, Taconic) or immunocompetent C57BL/6J mice (male, Jax strain # 000664). Mice were euthanized 12–14 days after injection.

Colony Formation Assays

One-thousand cells were plated in triplicate in 6-well plates and allowed to grow for two weeks before fixation in 10% formalin and staining with 0.5% crystal violet. Enzalutamide (10uM) or vehicle (DMSO) were added one day after plating. Colonies were quantified with ImageJ software (https://imagej.nih.gov/ij/) using the 'Analyze particle' tool with size 50-infinity and circularity 0.4 – 1.0. Assays were performed with a minimum of two independent biological replicates.

Statistical Analysis

Statistical analyses were performed using MAGECK-RRA (Robust Rank Analysis) [23] for CRISPRa/I screens and two-sample two-tailed t-test, one-way ANOVA with multiple comparison testing or Fisher's exact test as indicated in each figure legend. Kaplan–Meier survival analysis was conducted using GraphPad Prism software and analyzed with the log-rank test.

In box plots, the boxes show the 25th–75th percentile, center lines show the median and whiskers show the minimum–maximum values. In all bar graphs and dot-plots means are represented and error bars show the standard deviation unless specifically stated.

The sample size for mouse cohorts for library screens was determined a priori based on maintaining an estimated 1000X representation of each sgRNA library at initial inoculation of cells. For functional validation studies, sample size for intratibial and orthotopic injections was determined based on the maximum cage density allowable by IACUC (n=5 mice) for each condition, and no a priori assessment was made. Orthotopic experiments were performed in duplicate. No randomization was performed, and investigators were not blinded to experimental groups. GraphPad Prism v9 or R v.3.6.3 and R-studio 1.3.959–1 were used for statistical calculations and data visualization. In all graphs, asterisks were used to indicate significant P values according to the following thresholds: * for $P<0.05$, ** for $P<0.01$, *** for $P<0.001$, and **** for $P<0.0001$.

Gene set enrichment analysis (GSEA) was performed with GenePattern software[46] using gene signatures preranked by Spearman's coefficient or T-statistic and the Hallmarks Gene Sets from MSigDB (https://www.gsea-msigdb.org/gsea/msigdb), where NES and P values were estimated using 1,000 gene permutations, with FDR $p<0.1$ considered significant. For metastasis and dormancy pathways, c2-c8 Gene sets were analyzed. Gene sets containing the word "metastasis" and both Up- and Down-regulated gene sets were selected for the metastasis category. Gene sets containing "quiescent" were selected for dormancy categories, and the dormancy signature from Kim et al [34] was further included.

Genome-wide correlation and survival analysis was performed using data downloaded directly from cBioportal from the TCGA firehose legacy and DFKZ [32] (primary tumors) and SU2C/PCF dream team (SU2C [9]) and FHCRC [7] (metastasis) datasets. A median cutoff was used to group samples into CITED2 or E2F1 Low or High groups.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The research was supported by NIH grants R01 CA193442, R01 CA173481, R01 CA183929 and P01 CA265768 to CAS, and UH3 EB025765, P41 EB027062 and R01 CA249799 to GVN. JMA was supported by a postdoctoral training grant from the Department of Defense (DOD) Prostate Cancer Research Program (W81XWH-15-1-0185), an Irving Institute/Clinical Trials Office Pilot Award funded by the National Center for Advancing Translational Sciences, NIH (UL1TR001873), the Dean's Precision Medicine Research Fellowship from the Irving Institute for Clinical and Translational Research at CUIMC (UL1TR001873) and a Prostate Cancer Foundation Young Investigator Award (20YOUN25). CAS is supported by the TJ Martell Foundation for Leukemia, Cancer and AIDS Research and The Prostate Cancer Foundation and is an American Cancer Society Research Professor. FP was supported by a DOD Early Investigator Research Award (W81XWH-22-1-0054). This research was funded in part through the National Institutes of Health (NIH)/NCI Cancer Center Support Grant P30CA013696 awarded to the Herbert Irving Comprehensive Cancer Center (HICCC), which supported the Molecular Pathology, Flow Cytometry, Genomics and High Throughput Screening and Oncology Precision Therapeutics and Imaging Cores at HICCC. We are grateful to Jonathan Weissman for kindly providing the genome-wide libraries used in this study. Some figure panels (as indicated) were created with BioRender.com using an institutional license sponsored by Columbia University's VP&S Office for Research.

DATA AVAILABILITY STATEMENT:

Sequencing raw counts and differentially expressed sgRNAs are included in the manuscript tables. Plasmids and sgRNA libraries are available from Addgene. Code used for analysis

is available at https://sourceforge.net/p/mageck/wiki/Home/. RNAseq data was deposited in GEO, accession number GSE253815.

REFERENCES

1. Hernandez RK, Wade SW, Reich A, Pirolli M, Liede A, Lyman GH. Incidence of bone metastases in patients with solid tumors: analysis of oncology electronic medical records in the United States. BMC Cancer 2018; 18: 44. [PubMed: 29306325]
2. Gandaglia G, Abdollah F, Schiffmann J, Trudeau V, Shariat SF, Kim SP et al. Distribution of metastatic sites in patients with prostate cancer: A population-based analysis. Prostate 2014; 74: 210–216. [PubMed: 24132735]
3. Arriaga JM, Abate-Shen C. Genetically Engineered Mouse Models of Prostate Cancer in the Postgenomic Era. Cold Spring Harb Perspect Med 2018.
4. Grabowska MM, DeGraff DJ, Yu X, Jin RJ, Chen Z, Borowsky AD et al. Mouse models of prostate cancer: picking the best model for the question. Cancer Metastasis Rev 2014; 33: 377–397. [PubMed: 24452759]
5. Cancer Genome Atlas Research N. The Molecular Taxonomy of Primary Prostate Cancer. Cell 2015; 163: 1011–1025. [PubMed: 26544944]
6. Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature 2012; 487: 239–243. [PubMed: 22722839]
7. Kumar A, Coleman I, Morrissey C, Zhang X, True LD, Gulati R et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nat Med 2016; 22: 369–378. [PubMed: 26928463]
8. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS et al. Integrative genomic profiling of human prostate cancer. Cancer Cell 2010; 18: 11–22. [PubMed: 20579941]
9. Abida W, Cyrta J, Heller G, Prandi D, Armenia J, Coleman I et al. Genomic correlates of clinical outcome in advanced prostate cancer. Proc Natl Acad Sci U S A 2019; 116: 11428–11436. [PubMed: 31061129]
10. Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E et al. The long tail of oncogenic drivers in prostate cancer. Nat Genet 2018; 50: 645–651. [PubMed: 29610475]
11. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell 2014; 159: 647–661. [PubMed: 25307932]
12. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 2013; 154: 442–451. [PubMed: 23849981]
13. Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. Cell 2015; 160: 1246–1260. [PubMed: 25748654]
14. Bajaj J, Hamilton M, Shima Y, Chambers K, Spinler K, Van Nostrand EL et al. An in vivo genome-wide CRISPR screen identifies the RNA-binding protein Staufen2 as a key regulator of myeloid leukemia. Nat Cancer 2020; 1: 410–422. [PubMed: 34109316]
15. Dong MB, Wang G, Chow RD, Ye L, Zhu L, Dai X et al. Systematic Immunotherapy Target Discovery Using Genome-Scale In Vivo CRISPR Screens in CD8 T Cells. Cell 2019; 178: 1189–1204 e1123. [PubMed: 31442407]
16. Dai M, Yan G, Wang N, Daliah G, Edick AM, Poulet S et al. In vivo genome-wide CRISPR screen reveals breast cancer vulnerabilities and synergistic mTOR/Hippo targeted combination therapy. Nat Commun 2021; 12: 3055. [PubMed: 34031411]
17. Yau EH, Kummetha IR, Lichinchi G, Tang R, Zhang Y, Rana TM. Genome-Wide CRISPR Screen for Essential Cell Growth Mediators in Mutant KRAS Colorectal Cancers. Cancer Res 2017; 77: 6330–6339. [PubMed: 28954733]
18. Cheng L, Sun J, Pretlow TG, Culp J, Yang NS. CWR22 xenograft as an ex vivo human tumor model for prostate cancer gene therapy. J Natl Cancer Inst 1996; 88: 607–611. [PubMed: 8609662]

19. Sramkoski RM, Pretlow TG 2nd, Giaconia JM, Pretlow TP, Schwartz S, Sy MS et al. A new human prostate carcinoma cell line, 22Rv1. In Vitro Cell Dev Biol Anim 1999; 35: 403–409. [PubMed: 10462204]
20. Kovar JL, Johnson MA, Volcheck WM, Chen J, Simpson MA. Hyaluronidase expression induces prostate tumor metastasis in an orthotopic mouse model. Am J Pathol 2006; 169: 1415–1426. [PubMed: 17003496]
21. Preston Campbell J, Mulcrone P, Masood SK, Karolak M, Merkel A, Hebron K et al. TRIzol and Alu qPCR-based quantification of metastatic seeding within the skeleton. Scientific Reports 2015; 5: 12635. [PubMed: 26271202]
22. Horlbeck MA, Gilbert LA, Villalta JE, Adamson B, Pak RA, Chen Y et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. Elife 2016; 5.
23. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol 2014; 15: 554. [PubMed: 25476604]
24. Ronaldson-Bouchard K, Teles D, Yeager K, Tavakol DN, Zhao Y, Chramiec A et al. A multi-organ chip with matured tissue niches linked by vascular flow. Nat Biomed Eng 2022; 6: 351–371. [PubMed: 35478225]
25. Zou M, Toivanen R, Mitrofanova A, Floch N, Hayati S, Sun Y et al. Transdifferentiation as a Mechanism of Treatment Resistance in a Mouse Model of Castration-Resistant Prostate Cancer. Cancer Discov 2017; 7: 736–749. [PubMed: 28411207]
26. Arriaga JM, Panja S, Alshalalfa M, Zhao J, Zou M, Giacobbe A et al. A MYC and RAS co-activation signature in localized prostate cancer drives bone metastasis and castration resistance. Nat Cancer 2020; 1: 1082–1096. [PubMed: 34085047]
27. Shin SH, Lee GY, Lee M, Kang J, Shin HW, Chun YS et al. Aberrant expression of CITED2 promotes prostate cancer metastasis by activating the nucleolin-AKT pathway. Nat Commun 2018; 9: 4113. [PubMed: 30291252]
28. Bhattacharya S, Michels CL, Leung MK, Arany ZP, Kung AL, Livingston DM. Functional role of p35srj, a novel p300/CBP binding protein, during transactivation by HIF-1. Genes Dev 1999; 13: 64–75. [PubMed: 9887100]
29. Chou YT, Hsieh CH, Chiou SH, Hsu CF, Kao YR, Lee CC et al. CITED2 functions as a molecular switch of cytokine-induced proliferation and quiescence. Cell Death Differ 2012; 19: 2015–2028. [PubMed: 22814619]
30. Hu C, Zhang Y, Tang K, Luo Y, Liu Y, Chen W. Downregulation of CITED2 contributes to TGFbeta-mediated senescence of tendon-derived stem cells. Cell Tissue Res 2017; 368: 93–104. [PubMed: 28084522]
31. Mattes K, Berger G, Geugien M, Vellenga E, Schepers H. CITED2 affects leukemic cell survival by interfering with p53 activation. Cell Death Dis 2017; 8: e3132. [PubMed: 29072699]
32. Gerhauser C, Favero F, Risch T, Simon R, Feuerbach L, Assenov Y et al. Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. Cancer Cell 2018; 34: 996–1011 e1018. [PubMed: 30537516]
33. Pettaway CA, Pathak S, Greene G, Ramirez E, Wilson MR, Killion JJ et al. Selection of highly metastatic variants of different human prostatic carcinomas using orthotopic implantation in nude mice. Clin Cancer Res 1996; 2: 1627–1636. [PubMed: 9816342]
34. Kim RS, Avivar-Valderas A, Estrada Y, Bragado P, Sosa MS, Aguirre-Ghiso JA et al. Dormancy signatures and metastasis in estrogen receptor positive and negative breast cancer. PLoS One 2012; 7: e35569.
35. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020; 70: 7–30. [PubMed: 31912902]
36. Hubbard GK, Mutton LN, Khalili M, McMullin RP, Hicks JL, Bianchi-Frias D et al. Combined MYC Activation and Pten Loss Are Sufficient to Create Genomic Instability and Lethal Metastatic Prostate Cancer. Cancer Res 2016; 76: 283–292. [PubMed: 26554830]
37. Joyce JA, Pollard JW. Microenvironmental regulation of metastasis. Nat Rev Cancer 2009; 9: 239–252. [PubMed: 19279573]

38. Borriello L, Karagiannis GS, Duran CL, Coste A, Oktay MH, Entenberg D et al. The role of the tumor microenvironment in tumor cell intravasation and dissemination. Eur J Cell Biol 2020; 99: 151098.

39. Jayaraman S, Doucet M, Lau WM, Kominsky SL. CITED2 Modulates Breast Cancer Metastatic Ability through Effects on IKKalpha. Mol Cancer Res 2016; 14: 730–739. [PubMed: 27216153]

40. Lau WM, Weber KL, Doucet M, Chou YT, Brady K, Kowalski J et al. Identification of prospective factors promoting osteotropism in breast cancer: a potential role for CITED2. Int J Cancer 2010; 126: 876–884. [PubMed: 19642106]

41. Valastyan S, Weinberg RA. Tumor metastasis: molecular insights and evolving paradigms. Cell 2011; 147: 275–292. [PubMed: 22000009]

42. Lawson H, van de Lagemaat LN, Barile M, Tavosanis A, Durko J, Villacreces A et al. CITED2 coordinates key hematopoietic regulatory pathways to maintain the HSC pool in both steady-state hematopoiesis and transplantation. Stem Cell Reports 2021; 16: 2784–2797. [PubMed: 34715054]

43. Wang H, Zhang W, Bado I, Zhang XH. Bone Tropism in Cancer Metastases. Cold Spring Harb Perspect Med 2020; 10.

44. Easterly ME, Foltz CJ, Paulus MJ. Body condition scoring: comparing newly trained scorers and micro-computed tomography imaging. Lab Anim (NY) 2001; 30: 46–49. [PubMed: 11385757]

45. Villasante A, Marturano-Kruik A, Robinson ST, Liu Z, Guo XE, Vunjak-Novakovic G. Tissue-Engineered Model of Human Osteolytic Bone Tumor. Tissue Eng Part C Methods 2017; 23: 98–107. [PubMed: 28068876]

46. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. Nat Genet 2006; 38: 500–501. [PubMed: 16642009]

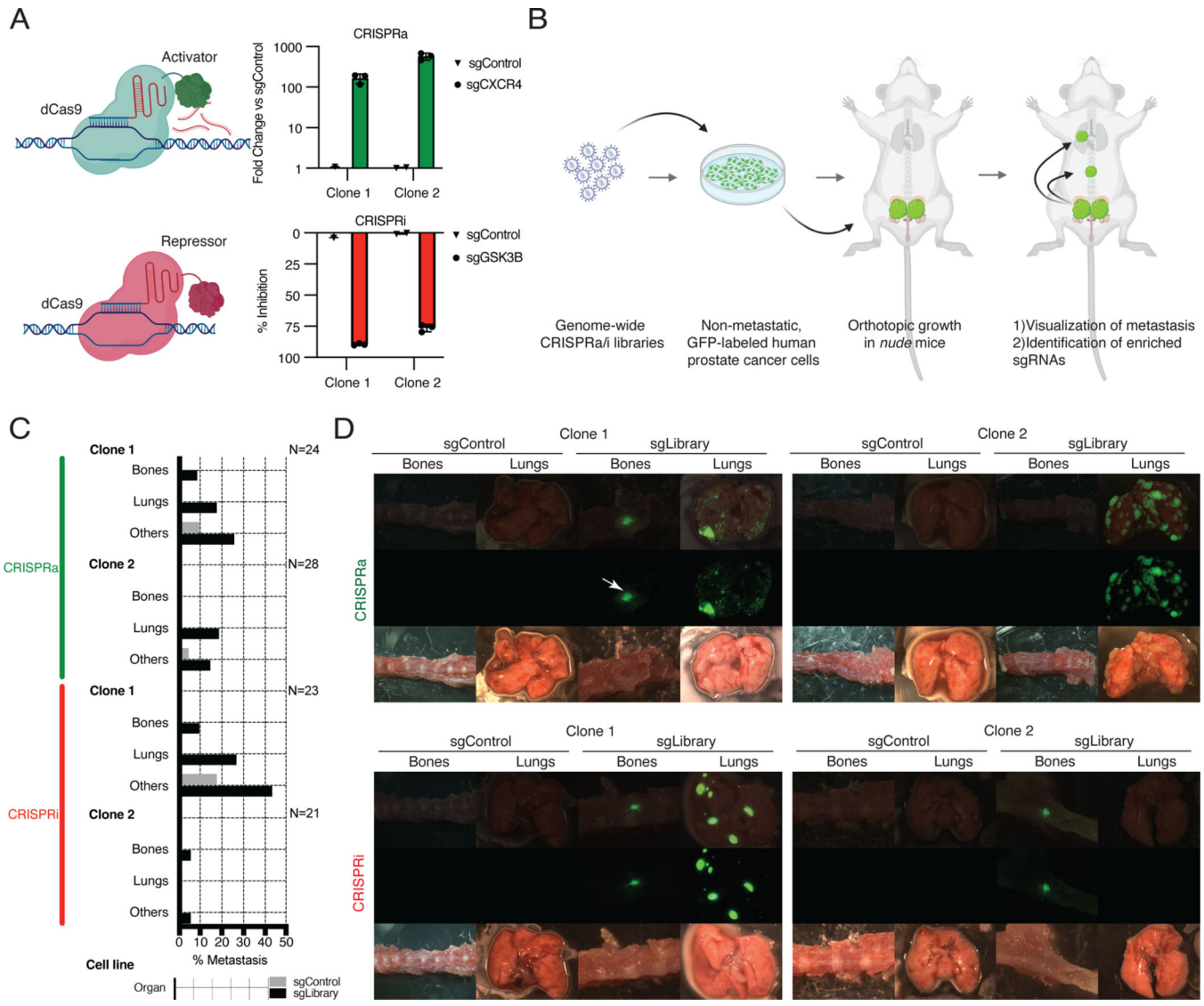

Figure 1:

In vivo genome-wide screening for activation and inhibition of gene expression results in metastasis of prostate cancer cells. (A) Engineering of 22Rv1 cell line derivatives to activate or inhibit gene expression of validated sgRNAs (sgCXCR4 and sgGSK3B, respectively), as measured by qRT-PCR. (B) Schematic outline of screen design. (C) Number and percentage of mice that developed metastasis following sgControl or sgLibrary infection of cell lines. (D) Representative *ex-vivo* bright-field (bottom row), epifluorescence (middle) and merged (top) images of bones and lungs showing presence or absence of GFP-labeled metastatic tumor cells after each screen. The arrow points to one example metastasis. Panels A and B were created with BioRender.com.

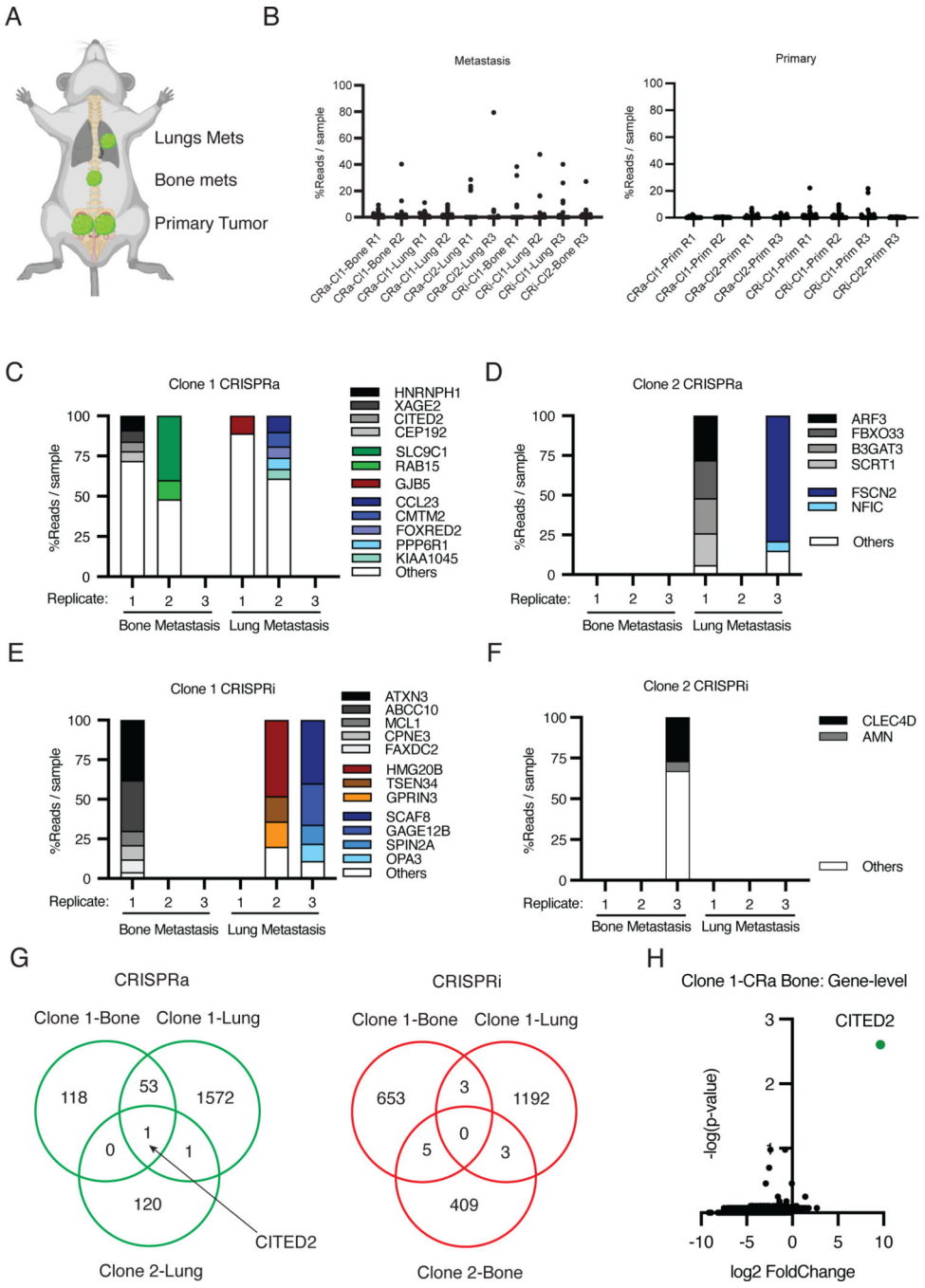

Figure 2:

Identification and analysis of metastasis hits in all genome-wide screens. (A) Schematic of samples sequenced to identify sgRNAs. (B) Scatter plots showing the representation of individual sgRNAs (shown as percentage of reads for each sgRNA over all reads in a sample) in metastasis (left) and their corresponding primary tumors (right). The dashed line shows the cutoff used for panels C-F. (C-F) Stacked bar graphs highlighting the identity of the “top-represented sgRNAs” (more than 5% of all reads in a particular metastatic sample) for each individual screen replicate. (G) Venn diagrams showing the overlap

between significantly enriched sgRNAs in each metastasis compared to primary tumors, after Mageck-RRA analysis (FDR-corrected P<0.001). An arrow highlights CITED2 as the only overlapping sgRNA in CRISPRa screens. (H) Volcano plot of gene-level Mageck-RRA analysis in bone metastasis of the 22Rv1- Clone 1 - CRa screen highlighting *CITED2* as the only significant hit (FDR-adjusted P=0.0025). Panel A was created with BioRender.com.

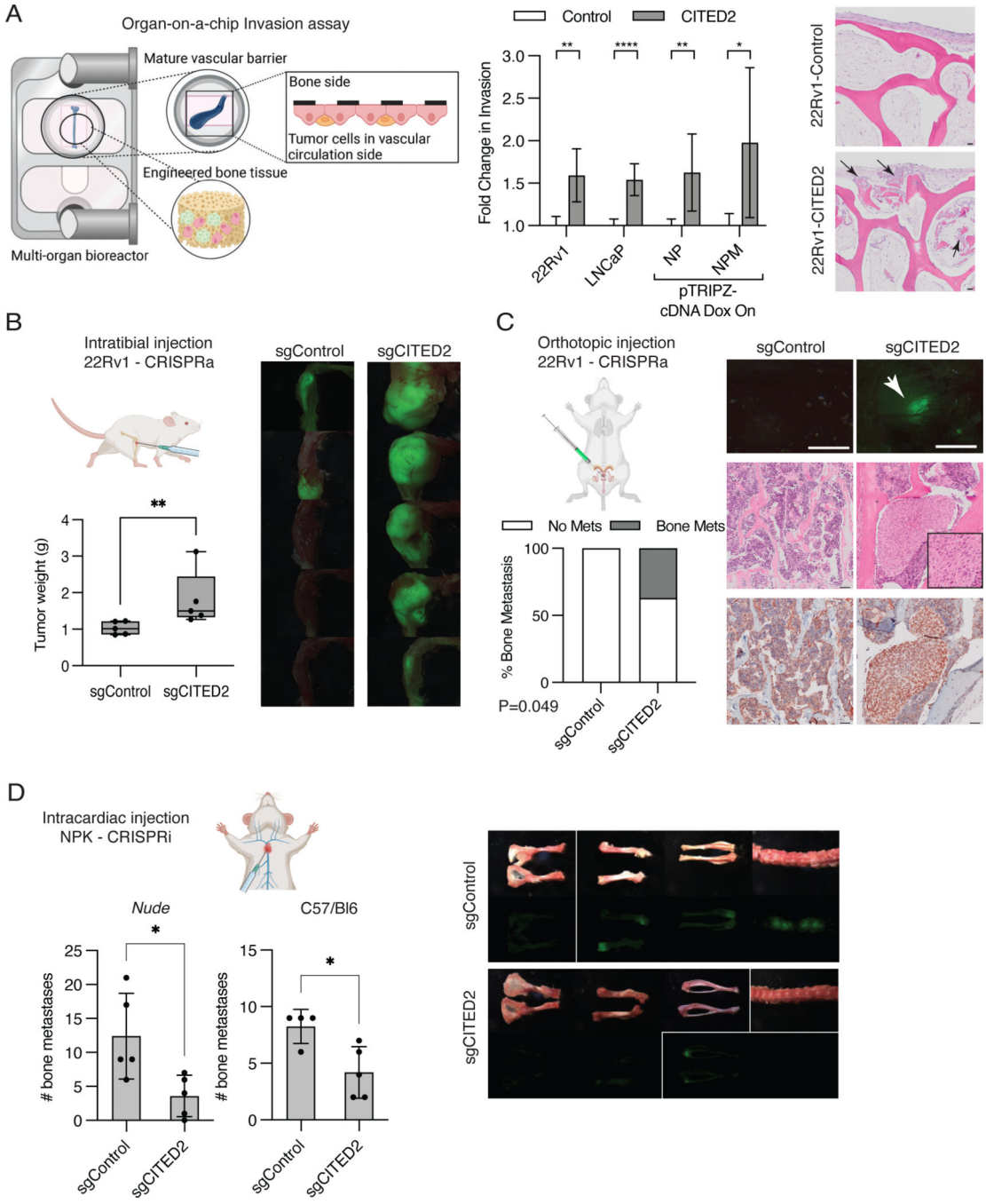

Figure 3:

Functional validation of CITED2 as a driver of prostate cancer bone metastasis. (A) Schematic of the organ-on-a-chip invasion assay. Bar graphs showing the relative invasion of CITED2-overexpressing cells compared to their corresponding controls after four weeks culture, as measured by fluorescence emission on the bone side of cultures. P<0.05 by two-tailed t-test. Representative H&E stains of invaded engineered bones after culture with 22Rv1-CRa cells infected with sgControl or sgCITED2. Example areas of invasion are shown with arrows. (B) Intratibial injection assays of 22Rv1-CRa cells infected with

sgCITED2 or sgControl. Quantification of final tumor weights, measured by whole tibiae weights. P<0.01 by two-tailed t-test. *Ex-vivo* epifluorescence images of injected tibiae showing GFP-expressing tumors. (C) Orthotopic injection assays of 22Rv1-CRa cells infected with sgCITED2 or sgControl. Stacked bar graphs showing the percentage of mice in which bone metastases were observed. P value is shown for two-tailed Fisher's exact test. Representative *ex-vivo* epifluorescence images of bones as well as H&E staining and CITED2 expression. (D) Intracardiac injection assays of bone metastatic NPK-CRi cells infected with sgControl or sgCITED2, into immunodeficient nude and immunocompetent C57/Bl6 mice (n=4 each). Scatter plot showing total numbers of bone metastasis observed per mouse. P<0.05 by two-tailed t-test. Representative *ex-vivo* brightfield (top) and epifluorescence (bottom) images of bones are shown. Scale bars represent 50um for histological sections and 0.1cm for epifluorescence images. Schematics were created with BioRender.com.

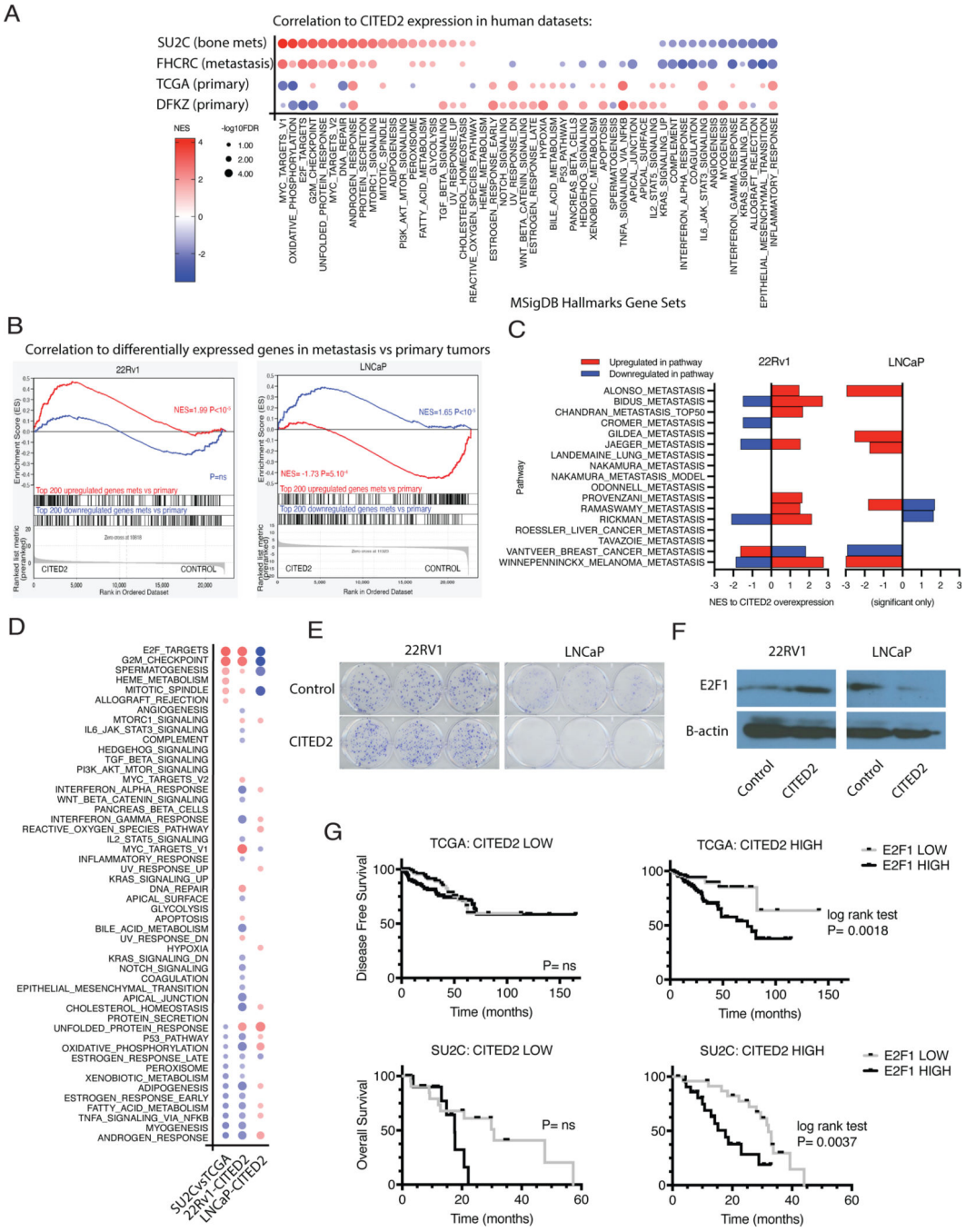

Figure 4:

Transcriptional outcome of CITED2 overexpression and relevance to human prostate cancer patients. (A) Bubble plots showing differentially enriched pathways upon inferred CITED2 transcriptional activity (genome-wide mRNA correlation with CITED2 mRNA levels) in different human prostate cancer datasets, shown in each row. The top two rows are metastasis and the bottom two are primary tumor datasets. Hallmark pathways from the Molecular Signatures Database (MSigDB) datasets available from the Broad Institute are shown. Bubbles are colored by NES and sized by FDR p-value as shown in the legend,

plotting only significant pathways (FDR <0.1). NES and P values were estimated using 1,000 gene permutations. (B) GSEA of genes preranked by differential expression upon CITED2 overexpression in 22Rv1 (left) and LNCaP (right), using as query the top 200 upregulated (red curve) or downregulated (blue curve) genes in bone metastases of the SU2C cohort (n=74) versus primary tumors of the TCGA cohort (n=497) (ie, 'SU2C vs TCGA'). (C) Stacked bar graphs showing metastasis pathway enrichment by GSEA upon CITED2 overexpression in 22Rv1 and LNCaP cells. All c2-c8 pathways with the term "metastasis" are shown. Only significant (FDR P<0.1) normalized enrichment scores (NES) are shown, with positive and negative values meaning positive and negative enrichment, respectively, upon CITED2 overexpression. Upregulated gene sets in each pathway are shown in red and downregulated gene sets in blue. (D) Bubble plots showing differentially enriched pathways upon SU2C vs TCGA (first column) and CITED2 overexpression in 22Rv1 (second column) and LNCaP cells (third column). Pathways, statistics, and scale as in (A). (E) Colony formation assays and (F) western blot analysis of E2F1 expression on 22Rv1 and LNCaP cells upon CITED2 overexpression. Quantification of colonies in (E) is shown in Figure S4H. (G). Kaplan-Meier survival curves of E2F1-low and E2F1-high patient groups. Graphs are shown for CITED2 -low (left) and CITED-high (right) cases in TCGA primary tumors (top) and in SU2C metastases (bottom).