JOURNAL OF
EMERGING INVESTIGATORS

Article

CDK7 inhibition disrupts androgen signaling and induces metabolic rewiring in prostate cancer cells

Brian Zhao[1], Erick Mitchell-Velasquez[2]
[1] Scarsdale High School, Scarsdale, New York
[2] University of Pennsylvania, Perelman School of Medicine, Philadelphia, Pennsylvania

SUMMARY

Prostate cancer (PC) is the second leading cause of cancer-related deaths among men in the United States. The lethality of advanced castration-resistant prostate cancer (CRPC), which is resistant to conventional androgen deprivation therapy, highlights the urgent need for new therapeutic strategies. In this study, we focused on how cyclin-dependent kinase 7 (CDK7) regulates transcriptional programs in PC. We hypothesized that inhibiting CDK7 would reprogram transcriptional networks in CRPC cells that may be associated with adaptive responses to treatment. To test this, we examined androgen receptor (AR) signaling pathways, which play major roles in PC proliferation. CDK7 phosphorylates Mediator complex subunit 1 (MED1), a coactivator of AR, enhancing the transcription of target genes that drive tumor growth. We assessed the effects of THZ1, a selective CDK7 inhibitor, on CRPC cell lines, including lymph node carcinoma of the prostate (LNCaP) and vertebral cancer of the prostate (VCaP). Through analysis of RNA-seq datasets, we found significant transcriptional changes following CDK7 inhibition, including downregulation of oncogenic pathways such as AR signaling and MYC targets, and notable upregulation of oxidative phosphorylation (OXPHOS). This suggests a metabolic shift that may help CRPC cells adapt to CDK7 inhibition. Gene set enrichment analysis highlighted the impact of CDK7 inhibition on metabolic and transcriptional pathways, underscoring the potential of targeting CDK7 in AR-driven PCs, including those resistant to androgen deprivation therapy. Combining CDK7 inhibitors with agents that target OXPHOS could improve therapeutic efficacy by simultaneously disrupting transcriptional and metabolic adaptations in treatment-resistant PC.

INTRODUCTION

Prostate cancer (PC) is one of the most commonly diagnosed diseases in men worldwide and is the second leading cause of cancer mortality in the male population (1). Androgen receptor (AR) signaling is strongly associated with tumor growth and aggressiveness in PC and, as such, is a target for androgen deprivation therapy, particularly in the context of castration-resistant prostate cancer (CRPC), a lethal form of the disease that emerges following resistance to androgen deprivation therapy (2). CRPC is characterized by the ability of the cancer cells to proliferate and metastasize even in the absence of circulating androgens, leading to poor patient outcomes and limited survival (2). The AR drives the expression of genes that promote PC cell growth and survival, and this process requires cooperation with transcriptional regulators. Mediator complex subunit 1 (MED1) acts as a coactivator of AR and is necessary for efficient transcription of AR target genes (4). Cyclin-dependent kinase 7 (CDK7), a central regulator of transcription, phosphorylates RNA polymerase II and modulates MED1 within the Mediator complex (5,6). This CDK7–MED1 axis enhances AR-driven transcriptional programs, including genes involved in proliferation, differentiation, and metabolism (6–8). Disruption of CDK7 activity impairs AR signaling and reduces expression of AR target genes, ultimately suppressing PC growth (4,9). These findings highlight CDK7 and MED1 as essential regulators of AR-dependent transcription and potential therapeutic targets in PC.

Recent advancements in molecular mechanisms have widened our understanding of the interactions between CDK7 and MED1, leading to the development of CDK7 inhibitors such as THZ1, which has shown great promise for PC treatment in preclinical models (4,5). These inhibitors exploit the dependence of certain cancer cells on high levels of transcription to sustain their uncontrolled growth, a phenomenon known as transcriptional addiction (10). However, it remains unclear how PC cells adapt to CDK7 inhibition and whether such adaptations contribute to therapeutic resistance. Our study addresses this gap by analyzing the transcriptional and metabolic reprogramming that occurs in CRPC cells after CDK7 inhibition, thereby providing insight into potential resistance mechanisms and informing strategies for more effective combination therapies.

Cancer cells undergo profound metabolic reprogramming to sustain proliferation and survive under therapeutic stress, and such adaptations are increasingly recognized as hallmarks of disease progression and resistance in PC (11–13). Transcriptional regulators like CDK7, through their control of RNA polymerase II and cofactors such as MED1, may influence not only oncogenic signaling but also the metabolic programs that support tumor growth (4–7). Based on this, we hypothesized that inhibiting CDK7 would reprogram transcriptional networks in CRPC cells, leading to alterations in metabolic pathways. To test this, we analyzed publicly available RNA-seq data from two CRPC cell lines: lymph node carcinoma of the prostate (LNCaP), which progresses to castration resistance, and vertebral cancer of the prostate (VCaP), a vertebral bone metastasis characterized by high AR expression and resistance to androgen deprivation therapy (2,4). Across both models, we identified commonly downregulated genes that highlight CDK7's role in regulating transcriptional programs, while upregulated genes point to adaptive shifts in cellular metabolism, including phosphatase

https://doi.org/10.59720/24-324

activity, oxidative stress responses, and mitochondrial function.

RESULTS

Experimental design and validation of CDK7 inhibition in CRPC cells

To determine how CDK7 inhibition affects transcriptional programs in CRPC cells, we analyzed publicly available RNA-seq data from LNCaP and VCaP cells treated with THZ1, a selective CDK7 inhibitor, as previously described (**Figure 1A**) (4). We exposed cells to 100 nmol/L THZ1 or vehicle control for 24 hours before RNA extraction. We performed principal component analysis (PCA) on the RNA-seq data and confirmed clear clustering of vehicle-treated and THZ1-treated samples along PC1 and PC2, validating distinct transcriptional shifts following CDK7 inhibition (**Figure 1B–C**). This clear separation of treated and control samples confirmed the robustness of the original experimental design and provided a strong foundation for our downstream analysis of CDK7-regulated gene expression changes (4). Preliminary differential expression analysis comparing THZ1-treated versus vehicle-treated samples within each cell line revealed extensive transcriptional changes in both LNCaP and VCaP cells. We visualized gene expression patterns using heatmaps to highlight treatment-associated expression shifts, and we used volcano plots to illustrate the distribution,

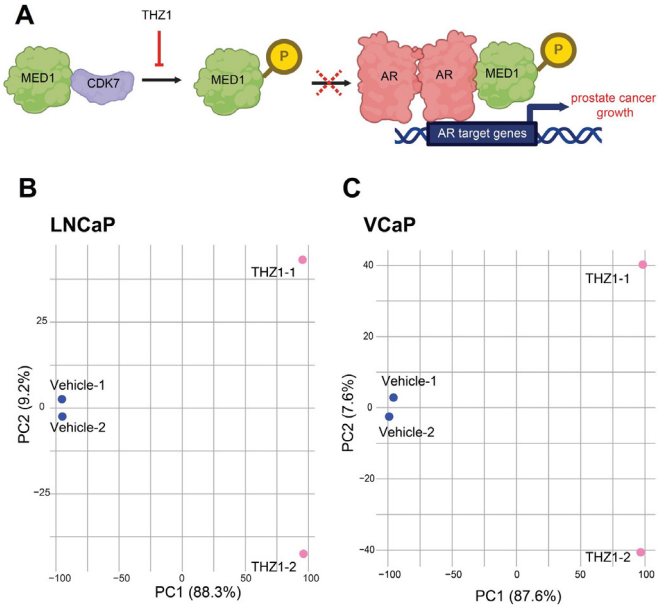

Figure 1: Experimental design and principal component analysis (PCA). A) Schematic showing the mechanism of mediator complex subunit 1 (MED1) phosphorylation by cyclin-dependent kinase 7 (CDK7) and its inhibition by THZ1. CDK7 phosphorylates MED1 at T1457, enhancing AR target gene transcription; THZ1 binds and inhibits CDK7, blocking this phosphorylation and downregulating AR signaling. **B).** PCA of lymph node carcinoma of the prostate (LNCaP) samples shows clear clustering of vehicle- versus THZ1-treated replicates. Principal component 1 explains 88.3% of the variance, and principal component 2 explains 9.2%. **C)** PCA of vertebral cancer of the prostate (VCaP) samples shows distinct clustering by treatment group, with principal component 1 explaining 87.6% of the variance and principal component 2 explaining 7.6%. Distinct transcriptional changes were seen following CDK7 inhibition in castration resistant prostate cancer (CRPC) cells.

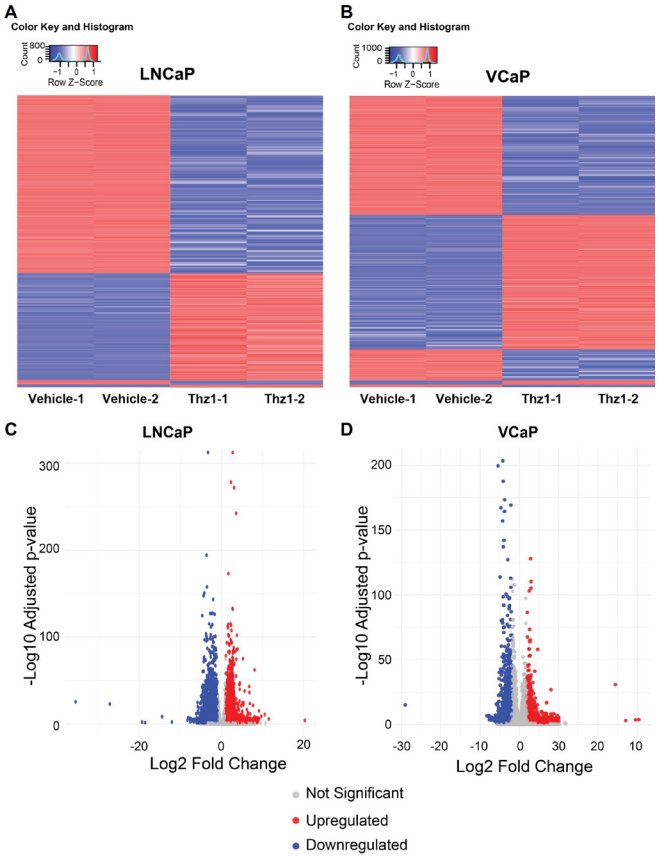

Figure 2: CDK7 inhibition alters global gene expression in CRPC cells. A-B) Heatmap of differentially expressed genes in **A)** LNCaP and **B)** VCap cells treated with 100 nM THZ1 or vehicle control for 24 hours. Volcano plot of differentially expressed genes in **C)** LNCaP and **D)** VCap cells treated with 100 nM THZ1 or vehicle control for 24 hours. Red indicates upregulated genes; blue indicates downregulated genes, scaled by row Z-score (n = 2 replicates per group).

magnitude, and statistical significance of these changes (**Figure 2**). These findings highlight CDK7's critical role in regulating transcriptional pathways in CRPC cell lines.

CDK7 inhibition alters metabolic and transcriptional pathways in PC

Applying statistical cutoffs of adjusted p-value < 0.05 and |log2 fold change| > 1.5 to the preliminary gene lists yielded a reduced set of significantly differentially expressed genes: 2,507 upregulated and 1,904 downregulated in LNCaP cells, and 1,498 upregulated and 1,435 downregulated in VCaP cells (**Figure 3**). Comparative analysis identified shared transcriptional responses to CDK7 inhibition across both cell lines. We found 714 genes that were upregulated and 621 genes that were downregulated in THZ1-treated cells compared to vehicle-treated controls in both LNCaP and VCaP lines (**Figure 3**).

To further investigate these changes, we performed gene ontology (GO) analysis on the commonly upregulated and downregulated genes across both cell lines. GO analysis of the upregulated genes revealed a strong enrichment for metabolic processes, particularly in catabolism and electron transport (**Figure 4A**). The top five GO terms identified

https://doi.org/10.59720/24-324

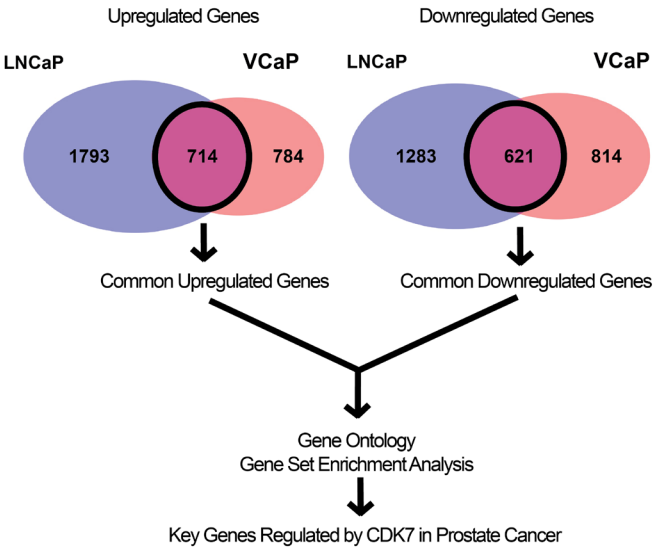

Figure 3: Overlap of transcriptional responses between LNCaP and VCaP cells. Venn diagrams illustrate the overlap of (left) upregulated and (right) downregulated genes in LNCaP and VCaP cells treated with 100 nM THZ1 versus vehicle control. Numbers represent the count of genes in each category. Shared gene sets were used for subsequent Gene Ontology (GO) and Gene Set Enrichment Analysis (GSEA).

were monocarboxylic acid catabolic process, organic acid catabolic process, carboxylic acid catabolic process, small molecule catabolic process, and electron transport chain. Conversely, GO analysis of the downregulated genes highlighted terms associated with transcriptional regulation and phosphatase activity (**Figure 4B**). Notable terms include DNA-binding transcription activator activity, DNA-binding transcription activator activity, RNA polymerase II-specific, protein tyrosine/threonine phosphatase activity, and MAP kinase tyrosine phosphatase activity. Overall, the identified pathways reflect transcriptional and metabolic changes associated with CDK7 inhibition in PC cells.

We then performed gene set enrichment analysis (GSEA) on the genes that were commonly up- or downregulated in both cell lines after THZ1 treatment. GSEA identified oxidative phosphorylation (OXPHOS) as a significantly upregulated hallmark pathway (**Figure 5**). Additionally, the reactive oxygen species (ROS) pathway was upregulated (**Figure 5**). Beyond metabolic alterations, CDK7 inhibition also led to the downregulation of genes involved in AR signaling, a key driver of PC progression (**Figure 5**). Genes associated with the transcription factors MYC and E2F were also downregulated (**Figure 5**). Collectively, these findings reveal alterations in metabolic and transcriptional pathways following CDK7 inhibition, including OXPHOS and androgen signaling pathways, in PC cells.

Expression of upregulated gene signatures in metastatic CRPC patients

To assess the clinical relevance of the adaptive programs we observed, we examined gene expression data from metastatic CRPC patients in the Stand Up to Cancer/Prostate Cancer Foundation (SU2C/PCF) Dream Team cohort using cBioPortal (14, 29, 30). We focused on the 26 gene sets and pathways that were upregulated following CDK7 inhibition in LNCaP and VCaP cells, as these signatures represent transcriptional programs that may become engaged in response to treatment. Among the upregulated pathways, the most prominent included oxidative phosphorylation, reactive oxygen species signaling, fatty acid metabolism, and epithelial–mesenchymal transition, which showed strong and consistent enrichment across both CRPC cell lines (**Figure 5**). To evaluate their relevance in patients, we queried the upregulated genes in this cohort and identified those expressed in at least 10% of metastatic CRPC tumors. We did not include downregulated pathways or GO terms in this comparison, as the aim was to highlight pathways induced following CDK7 inhibition. Notably, the pathways most frequently overexpressed in the patient cohort included

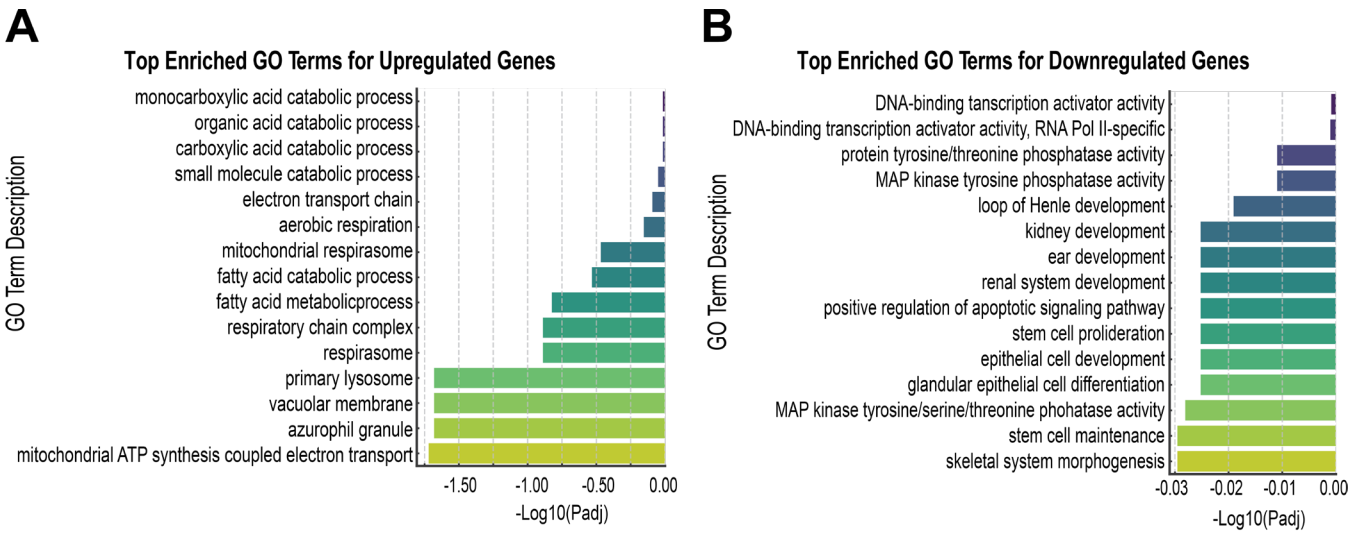

Figure 4: GO analysis shows CDK7 inhibition upregulates metabolic pathways and downregulates transcriptional and developmental programs in PC cells. A) Top enriched GO terms for upregulated genes in LNCaP and VCaP cells following THZ1 treatment (Padj < 0.05) highlight metabolic processes, including catabolism, electron transport, and mitochondrial respiration. **B)** Top enriched GO terms for downregulated genes reveal reduced expression of genes involved in transcription activator activity, phosphatase activity, and developmental pathways.

https://doi.org/10.59720/24-324

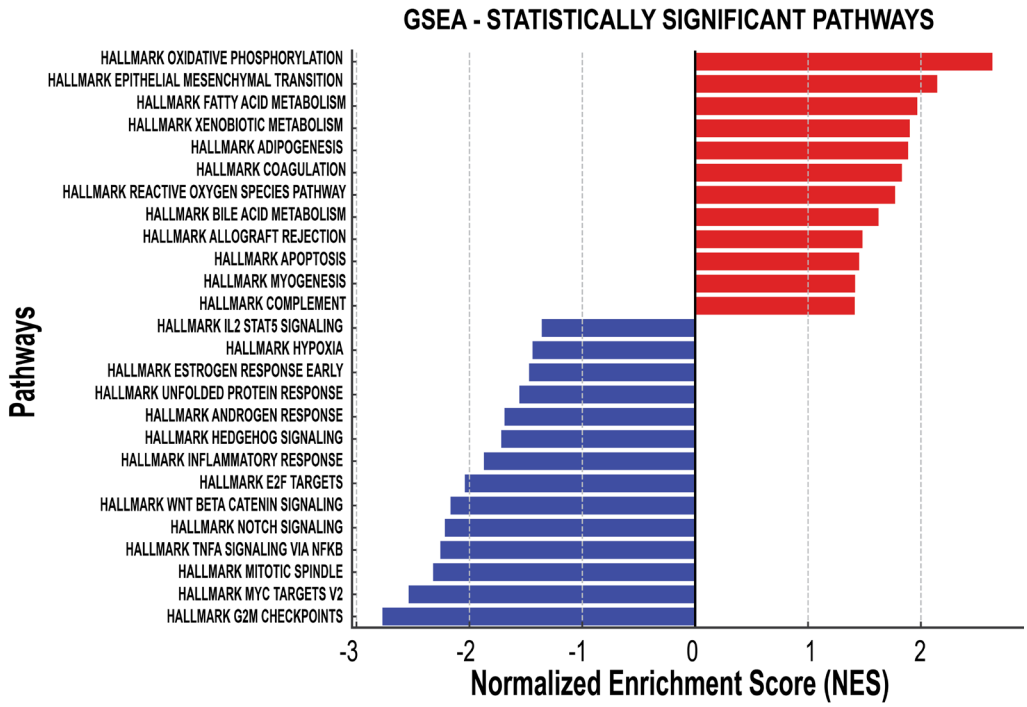

Figure 5: GSEA shows CDK7 inhibition upregulates metabolic pathways and downregulates cell cycle and proliferation pathways in PC cells. Statistically significant pathways enriched in THZ1-treated cells compared to vehicle control are shown. Upregulated pathways (red) include oxidative phosphorylation, epithelial-mesenchymal transition, and various metabolic processes. Downregulated pathways (blue) include G2/M checkpoint, MYC targets, E2F targets, and other proliferation-associated gene sets (adjusted p-value < 0.05).

OXPHOS, epithelial-mesenchymal transition, and fatty acid metabolism (**Figure 6**). These pathways are well-established drivers of cancer progression and drug resistance, highlighting their relevance in metastatic CRPC (11–13).

DISCUSSION

We hypothesized that inhibition of CDK7 would reprogram transcriptional networks in CRPC cells, leading to coordinated changes in metabolic pathways. Our analysis demonstrated that CDK7 inhibition produces widespread changes in gene expression programs in PC cells. By reanalyzing RNA-seq data from THZ1-treated CRPC cell lines, we found that CDK7 inhibition downregulates key transcriptional drivers such as *AR*, *MYC*, and *E2F* target genes, while upregulating metabolic pathways including OXPHOS. These findings highlight how disrupting CDK7 function reshapes both transcriptional and metabolic networks, providing insight into the adaptive responses of CRPC cells to transcriptional stress.

Our GSEA and GO analyses indicated that CDK7 inhibition was associated with downregulation of gene sets linked to AR signaling and MYC-driven transcriptional programs. These pathways are critical for PC cell survival, and their decreased representation in our analysis is consistent with CDK7's established role in regulating RNA polymerase II phosphorylation and transcriptional machinery (2, 5, 9, 15). AR-driven PC subtypes, including those resistant to conventional androgen deprivation therapy, may therefore be sensitive to disruption of CDK7-mediated transcription. This is supported by our findings in LNCaP and VCaP cells, both AR-positive CRPC models: LNCaP cells, originally androgen-sensitive, progress to castration resistance, while VCaP cells exhibit high AR expression and resistance to androgen deprivation (2,4). The transcriptional trends observed in these models suggest that CDK7 inhibition impairs AR-dependent programs, providing a rationale for exploring CDK7 inhibitors such as THZ1 in resistant disease contexts. While PC cells likely use additional mechanisms to suppress apoptosis, our analysis points to CDK7 as a potential regulatory vulnerability that warrants further experimental validation.

One of our key findings was that CDK7 inhibition was associated with upregulation of OXPHOS gene sets in CRPC cells. We highlight this pathway because it was the top hit in our GSEA analysis and represented the hallmark with the greatest number of genes recurrently expressed in metastatic CRPC patients. This shift in cellular metabolism may reflect an adaptive response, as CRPC cells increase mitochondrial energy production to compensate for the loss of CDK7-driven transcription. Similar metabolic adaptations have been reported in other cancers subjected to transcriptional stress or therapeutic pressure (16,17). The reliance on OXPHOS could help maintain ATP levels and support survival under these conditions (16,17). While complete inhibition of OXPHOS would broadly impair cell viability across many cell types, our findings suggest that CRPC cells may develop a heightened dependence on mitochondrial metabolism when CDK7 function is disrupted (11,16,17). This raises questions about the role of metabolic adaptation in therapeutic resistance and suggests that partial inhibition or combinatorial targeting of OXPHOS may enhance the efficacy of CDK7 inhibitors. Future studies using pharmacologic agents such as metformin or genetic approaches to attenuate, not abolish, OXPHOS activity (e.g., modulation of genes such as *COX6C, DECR1, IDH3G*, along with complementary metabolic profiling, could help determine how CDK7 inhibition reshapes cellular energy

https://doi.org/10.59720/24-324

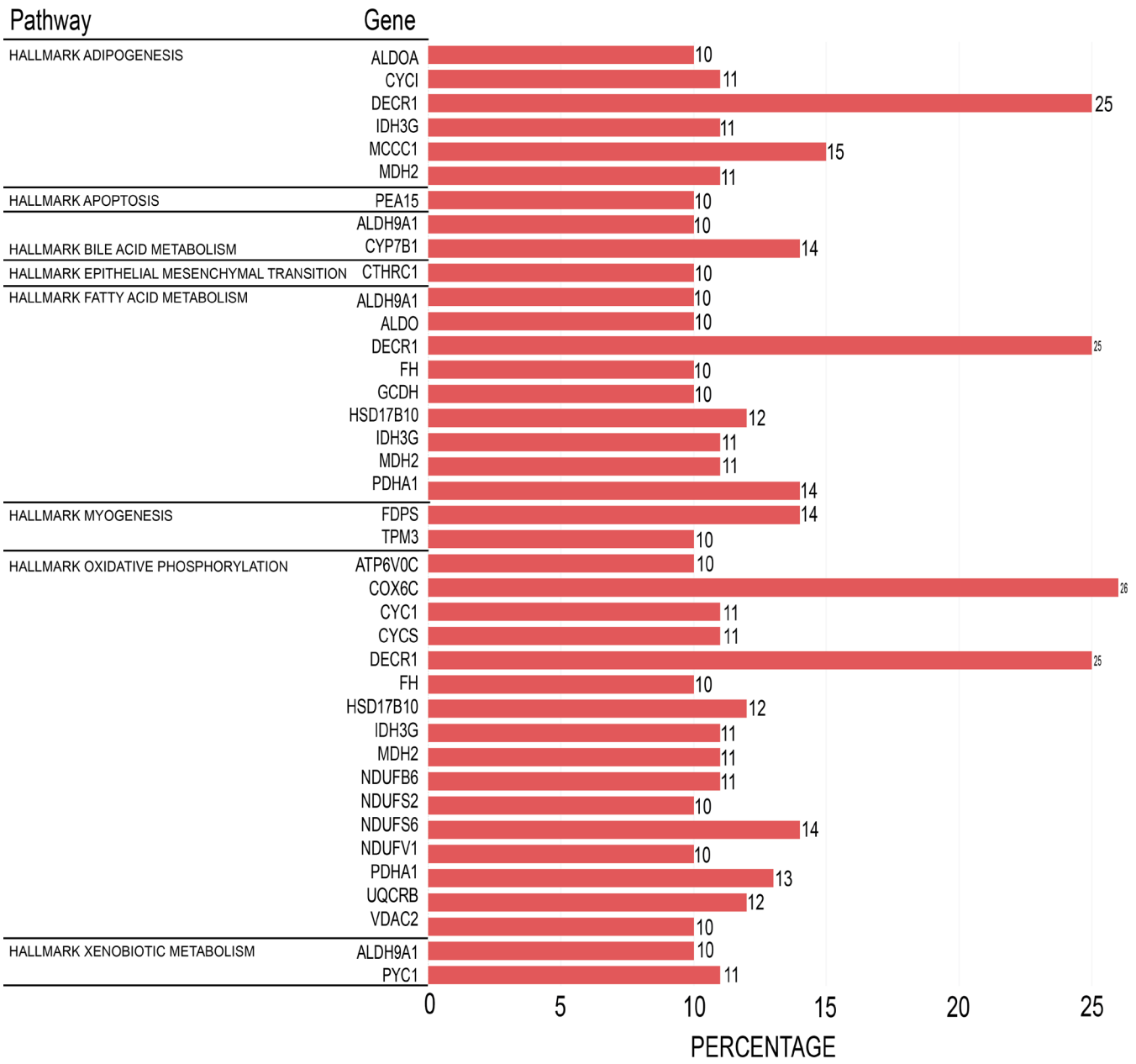

Figure 6: Genes upregulated by CDK7 inhibition show frequent overexpression in metastatic CRPC patients. Percentage of metastatic CRPC patients who overexpress genes upregulated following CDK7 inhibition in LNCaP and VCaP cells is shown. Patient expression data were obtained from the Stand Up to Cancer/Prostate Cancer Foundation (SU2C/PCF) Dream Team cohort (n = 118) via cBioPortal (14, 29, 30). These genes are involved in biological pathways commonly dysregulated in advanced PC, including oxidative phosphorylation, fatty acid metabolism, epithelial-mesenchymal transition, and apoptosis, highlighting their potential clinical significance for disease progression and therapeutic targeting.

homeostasis and whether OXPHOS represents a viable therapeutic vulnerability in CRPC.

A limitation of our study is that we relied on a single publicly available RNA-seq dataset of two CRPC cell lines, LNCaP and VCaP, which constrains the generalizability of our findings. Future analyses could expand to include additional PC models, such as androgen-sensitive versus androgen therapy-resistant cell lines, to better understand how CDK7 inhibition differentially affects these subtypes. Another limitation is that our conclusions are based on transcriptional analyses (RNA-seq, GSEA, and GO) and do not capture potential changes at the protein or functional level. Integrating complementary datasets, such as proteomic or metabolomic profiles, could strengthen the interpretation of CDK7-dependent regulatory networks and provide a more comprehensive picture of adaptive responses to CDK7 inhibition.

Altogether, our analysis shows that CDK7 inhibition reprograms gene expression in CRPC cells, downregulating AR and MYC-driven transcriptional networks while upregulating metabolic pathways such as OXPHOS. These findings highlight how transcriptional inhibition can reshape

https://doi.org/10.59720/24-324

tumor cell states and point to multiple avenues for future research, including the role of metabolic rewiring, the contribution of different PCa subtypes, and opportunities for rational combination therapies.

MATERIALS AND METHODS

Data acquisition and preprocessing

We obtained raw RNA-seq data from the Gene Expression Omnibus accession GSE125245 using the SRA Toolkit (version 3.0.0) (4, 18). Raw-transcript-level abundances were imported with tximport (19) All downstream analyses were performed in R (version 4.3.1) and RStudio (version 2023.09.2+464) (20, 21). Differential expression analysis was performed directly on raw counts using DESeq2 (version 1.34.0), which applies its internal size factor normalization and dispersion estimation procedures (22). After running DESeq2, we removed genes without adjusted p-values and filtered results for significance (adjusted p-value < 0.05 and |log2 fold change| > 1.5).

Principal component analysis (PCA)

To observe distinctions between gene expression data and reduce dimensionality, we performed PCA on normalized count data incorporated within the Seurat package (version 5.0.3), and the variance explained by each principal component was calculated (23). The first two components were visualized to display sample distribution, with samples labeled and color-coded according to their experimental groups.

Differential gene expression (DGE) analysis

We conducted DGE analysis using DESeq2 (version 1.34.0) to identify genes with significant changes in expression between experimental groups (22). Filtered data were input into DESeq2 to model gene expression as a function of treatment conditions. The model was fitted using the DESeq2 default method, which employs a negative binomial distribution and includes an internal normalization for library size. Gene expression differences between the vehicle and THZ1-treated samples (n=2 per group) were assessed using the Wald test. Adjustments for multiple testing were made using the Benjamini-Hochberg procedure. Genes with an adjusted p-value (P_{adj}) less than 0.05 and an absolute $\log_2$ fold change greater than one were considered significantly differentially expressed. A Venn diagram was constructed to visualize the overlap of differentially expressed genes (DEGs) between different experimental conditions, providing insights into shared and unique gene expression changes. Using the VennDiagram package (version 1.7.3), pairwise overlaps of significantly upregulated and downregulated genes ($|\log_2 FoldChange| > 1.5$, adjusted p-value < 0.05) were calculated between LNCaP and VCaP cells treated with the CDK7 inhibitor, and the diagrams were rendered with draw. pairwise.venn() for each category (24).

Hierarchical clustering and heatmap generation

To identify patterns of gene expression across samples, we performed hierarchical clustering on the DEGs. Heatmaps were generated using the pheatmap package (version 1.0.12), which applies hierarchical clustering to both genes (rows) and samples (columns) using Euclidean distance and complete linkage by default (25). This approach enabled us to uncover the underlying structure in the data and group genes with similar expression profiles. To enhance interpretability, we employed custom color palettes and annotation tracks, allowing for the clear visualization of expression patterns and gene clusters across experimental conditions.

Gene ontology (GO) analysis

To identify biological processes associated with CDK7 inhibition, we performed GO enrichment analysis on the common upregulated and downregulated genes in LNCaP and VCaP cells. GO analysis was conducted using the clusterProfiler package (verison 4.6.2) with the org.Hs.eg.db annotation database (version 3.17.0) (26, 27). Significantly enriched GO terms were defined as those with a Benjamini–Hochberg adjusted p-value < 0.05. Results were visualized as bar plots of the top enriched biological processes using ggplot2 (version3.4.4) (28).

Gene set enrichment analysis (GSEA)

We performed GSEA on the DEGs to identify biological pathways associated with gene expression changes under various experimental conditions. The analysis was conducted using clusterProfiler (version 4.6.2), applying the gseGO() function to assess enrichment of gene ontology biological process terms (26). To determine significantly enriched pathways, we applied an adjusted p-value (Padj) threshold of < 0.05 and a normalized enrichment score (NES) cutoff of |NES| ≥ 1.5. The results were visualized using the ggplot2 package (version 3.4.4), where bar plots were generated to display the NES values of statistically significant gene sets (28). Additionally, heatmaps summarizing NES values across datasets were created with pheatmap to highlight shared and unique pathway enrichments.

Gene expression assessment in CRPC patients

Gene expression profiles of 118 metastatic CRPC patients from the Stand Up to Cancer/Prostate Cancer Foundation (SU2C/PCF) Dream Team cohort were examined using cBioPortal (14, 29, 30). From our DEG analysis of LNCaP and VCaP cells treated with THZ1, we generated a list of genes that were consistently upregulated across both cell lines. These genes were then cross-referenced with patient tumor expression data in the SU2C/PCF cohort to assess their prevalence. Genes detected in at least 10% of patients were considered recurrently expressed and were subsequently plotted to illustrate their distribution within the metastatic CRPC population.

ACKNOWLEDGMENTS

We would like to thank Ying Xu for her support during this project.

Received: October 27, 2024
Accepted: May 12, 2025
Published: March 23, 2026

REFERENCES

1. Siegel, Rebecca L., et al. “Cancer statistics, 2024.” *CA: A Cancer Journal for Clinicians*, vol. 70, no. 1, 17 January 2024, pp.12-49. https://doi.org/10.3322/caac.21820
2. Watson, Philip A., et al. “Emerging Mechanisms of Resistance to Androgen Receptor Inhibitors in Prostate

https://doi.org/10.59720/24-324

Cancer." *Nature Reviews Cancer*, vol. 15, no. 12, 13 November 2015, pp. 701-711. https://doi.org/10.1038/nrc4016
3. Bolton, Eric C., et al. "Cell- and gene-specific regulation of primary target genes by the androgen receptor." *Genes & Development*, vol. 21, no. 16, 15 August 2007, pp. 2005-2017. https://doi.org/10.1101/gad.1564207
4. Rasool, Reyaz Ur, et al. "CDK7 Inhibition Suppresses Castration-Resistant Prostate Cancer through MED1 Inactivation." *Cancer Discovery*, vol. 9, no. 11, 1 November 2019, pp. 1538-1555. https://doi.org/10.1158/2159-8290.CD-19-0189
5. Kwiatkowski, Nicholas, et al. "Targeting Transcription Regulation in Cancer with a Covalent CDK7 Inhibitor." *Nature*, vol. 511, no. 7511, 22 June 2014, pp. 616-620. https://doi.org/10.1038/nature13393
6. Malik, Sohail and Robert G Roeder. "The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation." *Nature Reviews Genetics*, vol. 11, no. 11, 13 October 2010, pp. 761-72. https://doi.org/10.1038/nrg2901
7. Wang, Qianben, et al. "Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking." *Molecular Cell*, vol. 19, no. 5, 2 September 2005, pp. 631-42. https://doi.org/10.1016/j.molcel.2005.07.018
8. Schachter, Miriam M., et al. "A Cdk7-Cdk4 T-loop Phosphorylation Cascade Promotes G1 Progression." *Molecular Cell*, vol. 50, no. 2, 25 April 2013, pp. 250-260. https://doi.org/10.1016/j.molcel.2013.04.003
9. Chipumuro, Edmond, et al. "CDK7 Inhibition Suppresses Super-Enhancer-Linked Oncogenic Transcription in MYCN-Driven Cancer." *Cell*, vol. 159, no. 5, 20 November 2014, pp. 1126-1139. https://doi.org/10.1016/j.cell.2014.10.024
10. Zanconato, Francesca., et al. "Transcriptional addiction in cancer cells is mediated by YAP/TAZ through BRD4." *Nature Medicine*, vol 24, no. 10, 17 September 2018, pp. 1599-1610. https://doi.org/10.1038/s41591-018-0158-8
11. Martin, Matthew J., et al. "Inhibition of Oxidative Phosphorylation Suppresses the Development of Osimertinib Resistance in a Preclinical Model of EGFR-Driven Lung Adenocarcinoma." *Oncotarget*, vol. 7, no. 52, 16 November 2016, pp. 86313-86325. https://doi.org/10.18632/oncotarget.13388
12. Safa, Ahmad R. "Epithelial-Mesenchymal Transition: A Hallmark in Pancreatic Cancer Stem Cell Migration, Metastasis Formation, and Drug Resistance." *Journal of Cancer Metastasis and Treatment*, vol. 6, 26 Sept. 2020, p. 36. https://doi.org/10.20517/2394-4722.2020.55
13. Zadra, Giorgia, et al. "Inhibition of de Novo Lipogenesis Targets Androgen Receptor Signaling in Castration-Resistant Prostate Cancer." *Proceedings of the National Academy of Sciences*, vol. 116, no. 2, 21 December 2018, pp. 631-640. https://doi.org/10.1073/pnas.1808834116
14. Abida, Wassim et al. "Genomic correlates of clinical outcome in advanced prostate cancer." *Proceedings of the National Academy of Sciences,* vol. 116, no. 23, 6 May 2019, pp.11428-11436. https://doi.org/10.1073/pnas.1902651116
15. Fisher, Robert P. "Cdk7: a kinase at the core of transcription and in the crosshairs of cancer drug discovery." *Transcription,* vol. 10, no. 2, 6 December 2019, pp. 47-56. https://doi.org/10.1080/21541264.2018.1553483
16. Viale, Andrea, et al. "Oncogene ablation-resistant pancreatic cancer cells depend on mitochondrial function." *Nature,* vol. 514, no. 7524, 10 August 2014, pp. 628-632. https://doi.org/10.1038/nature13611
17. Ashton, Thomas M., et al. "Oxidative Phosphorylation as an Emerging Target in Cancer Therapy." *Clinical cancer research: an official journal of the American Association for Cancer Research*, vol. 24, no.11, 31 May 2018, pp. 2482-2490. https://doi.org/10.1158/1078-0432.ccr-17-3070
18. Leinonen, R., Sugawara, H., and Shumway, M. "The Sequence Read Archive." Nucleic Acids Research, vol. 39, suppl_1, 1 January 2011, pp. D19–D21. https://doi.org/10.1093/nar/gkq1019
19. Soneson, Charlotte, Michael I. Love, and Mark D. Robinson. "Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences." *F1000Research*, vol. 4, 28 December 2015, p. 1521. https://doi.org/10.12688/f1000research.7563.2
20. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2023. https://www.r-project.org/
21. RStudio: Integrated Development Environment for R., 2023. https://posit.co/products/open-source/rstudio/
22. Love, Michael I., et al. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology,* vol. 15, no. 12, 5 December 2014, p. 550. https://doi.org/10.1186/s13059-014-0550-8
23. Hao, Yuhan, et al. "Integrated analysis of multimodal single-cell data." *Cell* , vol. 184, no. 13, 24 June 2021, pp. 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048
24. Chen, Hanbo, and Paul K. Sharp. "VennDiagram: a package for the generation of highly customizable Venn and Euler diagrams in R." *BMC Bioinformatics,* vol. 12, 12 January 2011, p. 35. https://doi.org/10.1186/1471-2105-12-35
25. Kolde, Raivo. pheatmap: Pretty Heatmaps. R package version 1.0.12, 2019. https://cran.r-project.org/package=pheatmap
26. Yu, Guangchuang, et al. "clusterProfiler: an R package for comparing biological themes among gene clusters." *OMICS: A Journal of Integrative Biology*, vol. 16, no. 5, May 2012, pp. 284–287. https://doi.org/10.1089/omi.2011.0118
27. Carlson, Marc.org.Hs.eg.db: Genome Wide Annotation for Human. R package version 3.17.0, 2023. https://bioconductor.org/packages/org.Hs.eg.db
28. Wickham, Hadley, et al. "ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics." R package version 3.4.4, 2023. https://CRAN.R-project.org/package=ggplot2
29. Cerami, Ethan, et al. "The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data." *Cancer Discovery*, vol. 2, no. 5, May 2012, pp. 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095
30. Gao, Jianjiong, et al. "Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the

https://doi.org/10.59720/24-324

cBioPortal." *Science Signaling,* vol. 6, no. 269, 2 April 2013, pl1. https://doi.org/10.1126/scisignal.2004088

APPENDIX

```
# Load libraries
library(tximport)
library(ensembldb)
library(EnsDb.Hsapiens.v86)
library(DESeq2)
library(ggplot2)
library(clusterProfiler)
library(org.Hs.eg.db)
library(VennDiagram)

# Set file paths
data_dir <- "path/to/data"
sample_file <- file.path(data_dir, "sample_info.txt")
count_files <- file.path(data_dir, "sample_counts/")

# Load sample information
samples <- read.table(sample_file, header = TRUE, stringsAsFactor = FALSE)

# Load counts
txi_kallisto <- tximport(count_files, type = "kallisto", tx2gene = Tx, reader = read_tsv)

# Create DESeq2 dataset
dds <- DESeqDataSetFromTximport(txi_kallisto, colData = samples, design = ~ condition)

# Data preprocessing
dds <- estimateSizeFactors(dds)
dds <- estimateDispersions(dds)

# Run differential expression
dds <- DESeq(dds)

# Extract results
res <- results(dds, contrast = c("condition", "treatment", "control"))

# Adjust for multiple testing
res <- res[!is.na(res$padj), ]
res$padj <- p.adjust(res$pvalue, method = "BH")

# Filter significant results
sig_res <- res[res$padj < 0.05 & abs(res$log2FoldChange) > 1.5, ]

# Plotting
ggplot(sig_res, aes(x = log2FoldChange, y = -log10(padj))) +
  geom_point(aes(color = padj < 0.05)) +
  scale_color_manual(values = c("red", "black")) +
  labs(title = "Volcano plot of differentially expressed genes")

# Save results
write.csv(results_LNCaP, file = "results_LNCaP.csv")
write.csv(results_VCaP, file = "results_VCaP.csv")

# UMAP Visualization (Assuming PCA has been performed)
library(Seurat)
seurat_obj <- CreateSeuratObject(counts = dds@assays$counts)
seurat_obj <- NormalizeData(seurat_obj)
seurat_obj <- FindVariableFeatures(seurat_obj)
seurat_obj <- ScaleData(seurat_obj)
seurat_obj <- RunPCA(seurat_obj)
seurat_obj <- RunUMAP(seurat_obj, dims = 1:10)

DimPlot(seurat_obj, reduction = "umap", group.by = "condition")

# Load data
dds_LNCaP <- read.csv("results_LNCaP.csv", row.names = 1)
dds_VCaP <- read.csv("results_VCaP.csv", row.names = 1)

# Differential expression analysis
dds_LNCaP <- DESeq(dds_LNCaP)
dds_VCaP <- DESeq(dds_VCaP)

# Extract results
results_LNCaP <- results(dds_LNCaP, contrast = c("condition", "THZ1", "Vehicle"))
results_VCaP <- results(dds_VCaP, contrast = c("condition", "THZ1", "Vehicle"))

# Volcano Plot Function
plot_volcano <- function(res, title) {
  ggplot(res, aes(x=log2FoldChange, y=-log10(pvalue))) +
    geom_point(aes(color = padj < 0.05)) +
    scale_color_manual(values = c("grey", "red")) +
      labs(title = title, x = "Log2 Fold Change", y = "-Log10 p-value") +
    theme_minimal()
}

volcano_LNCaP <- plot_volcano(results_LNCaP, "Volcano Plot - LNCaP")
volcano_VCaP <- plot_volcano(results_VCaP, "Volcano Plot - VCaP")
print(volcano_LNCaP)
print(volcano_VCaP)

# Identify differentially expressed genes
significant_LNCaP <- subset(results_LNCaP, padj < 0.05 & abs(log2FoldChange) > 1.5)
significant_VCaP <- subset(results_VCaP, padj < 0.05 & abs(log2FoldChange) > 1.5)
```

https://doi.org/10.59720/24-324

```
# List of upregulated and downregulated genes
upregulated_LNCaP <- significant_LNCaP$gene[significant_LNCaP$log2FoldChange > 1.5]
upregulated_VCaP <- significant_VCaP$gene[significant_VCaP$log2FoldChange > 1.5]
downregulated_LNCaP <- significant_LNCaP$gene[significant_LNCaP$log2FoldChange < -1.5]
downregulated_VCaP <- significant_VCaP$gene[significant_VCaP$log2FoldChange < -1.5]

# Find common genes
common_upregulated <- intersect(upregulated_LNCaP, upregulated_VCaP)
common_downregulated <- intersect(downregulated_LNCaP, downregulated_VCaP)

# Save overlapping genes
write.csv(common_upregulated, "common_upregulated_genes.csv", row.names = FALSE)
write.csv(common_downregulated, "common_downregulated_genes.csv", row.names = FALSE)

# Create Venn diagram for upregulated genes
venn.plot_up <- draw.pairwise.venn(
  area1 = length(upregulated_LNCaP),
  area2 = length(upregulated_VCaP),
  cross.area = length(common_upregulated),
  category = c("LNCaP Up", "VCaP Up"),
  fill = c("blue", "red"),
  lty = "blank",
  cex = 2,
  cat.cex = 2,
  cat.pos = c(-20, 20),
  cat.dist = 0.05,
  main = "Overlap of Upregulated Genes"
)

# Create Venn diagram for downregulated genes
venn.plot_down <- draw.pairwise.venn(
  area1 = length(downregulated_LNCaP),
  area2 = length(downregulated_VCaP),
  cross.area = length(common_downregulated),
  category = c("LNCaP Down", "VCaP Down"),
  fill = c("green", "orange"),
  lty = "blank",
  cex = 2,
  cat.cex = 2,
  cat.pos = c(-20, 20),
  cat.dist = 0.05,
  main = "Overlap of Downregulated Genes"
)

# Display the plots
grid.arrange(venn.plot_up, venn.plot_down, ncol = 2)

# Gene Ontology and GSEA Analysis
ego_up <- enrichGO(gene = common_upregulated,
                   OrgDb = org.Hs.eg.db,
                   keyType = "SYMBOL",
                   ont = "BP",
                   pAdjustMethod = "BH",
                   qvalueCutoff = 0.05)

ego_down <- enrichGO(gene = common_downregulated,
                     OrgDb = org.Hs.eg.db,
                     keyType = "SYMBOL",
                     ont = "BP",
                     pAdjustMethod = "BH",
                     qvalueCutoff = 0.05)

# GSEA analysis
gsea_result_up <- gseGO(geneList = significant_LNCaP$log2FoldChange,
                        OrgDb = org.Hs.eg.db,
                        ont = "BP",
                        nPerm = 1000,
                        minGSSize = 10,
                        maxGSSize = 500,
                        pAdjustMethod = "BH",
                        pvalueCutoff = 0.05)

gsea_result_down <- gseGO(geneList = significant_VCaP$log2FoldChange,
                          OrgDb = org.Hs.eg.db,
                          ont = "BP",
                          nPerm = 1000,
                          minGSSize = 10,
                          maxGSSize = 500,
                          pAdjustMethod = "BH",
                          pvalueCutoff = 0.05)

# Save GSEA results
write.csv(gsea_result_up@result, "GSEA_Upregulated.csv")
write.csv(gsea_result_down@result, "GSEA_Downregulated.csv")

# Graphing GSEA results
gsea_combined <- rbind(
  transform(gsea_result_up@result, Significance = "Upregulated"),
  transform(gsea_result_down@result, Significance = "Downregulated")
)

# Filter results to show only significant pathways if needed (assuming using qvalue or adjusted p-values)
gsea_filtered <- subset(gsea_combined, padj < 0.05)

# Creating the plot
ggplot(gsea_filtered, aes(x = reorder(ID, -NES), y = NES, fill = Significance)) +
  geom_bar(stat = "identity") +
  coord_flip() + # Flip axes for horizontal bars
  scale_fill_manual(values = c("Upregulated" = "red", "Downregulated" = "blue")) +
  labs(title = "GSEA - Statistically Significant Pathways",
       x = "Pathway",
       y = "Normalized Enrichment Score (NES)") +
  theme_minimal() +
  theme(legend.position = "top", legend.title = element_blank())
```