Supplementary Materials

Functional deficiency of DNA repair gene EXO5 results in androgen-induced genomic instability and prostate tumorigenesis

Shafat Ali[1], Yilan Zhang[1], Mian Zhou[1], Hongzhi Li[2], Weiwei Jin[1], Li Zheng[1], Xiaochun Yu[1], Jeremy M. Stark[1], Jeffrey N. Weitzel[3], and Binghui Shen[1]*,

[1]Department of Cancer Genetics and Epigenetics, [2]Department of Molecular Medicine, and [3]Department of Population Sciences, Beckman Research Institute of City of Hope, 1500 East Duarte Rd, Duarte, CA, 91010, USA

Supplementary methods

Cell lines and reagents

The LNCaP, LAPC4, VCap, and derivative cell lines, including LNCaP-HDR, LNCaP-NHEJ, *EXO5* knockout, and *EXO1* knockout cells, were maintained in RPMI1640 media supplemented with 10% fetal bovine serum (Thermo Fisher Scientific) and 1% penicillin-streptomycin. The HEK293 cell line was maintained in DMEM supplemented with 10% fetal bovine albumin and 1% penicillin-streptomycin. All cell lines were obtained from American Type Culture Collection and were grown at 37^{o}C in 5% CO_2. We generated homozygous *EXO5* and *EXO1* knockout cell lines using the CRISPR-Cas9 system. The LNCaP cells harbored heterozygous *EXO5* L151P mutations. We designed at least two different guide RNA sequences complementary to the coding sequence of the *EXO5* and *EXO1*genes, based on its protospacer-adjacent motif (PAM) sequence (Supplementary Table 3). These guide sequences were cloned into a Px459 vector containing the Cas9 coding gene and a puromycin selection marker. LNCaP cells were then transiently transfected using Lipofectamine 2000 reagent (Life Technologies) and selected for puromycin resistance. Single-cell cloning was performed to isolate independent clones. Knockout of the gene was confirmed by direct Sanger sequencing.

Dihydrotestosterone (DHT), VP-16, and FLAG-M2 magnetic beads were purchased from Sigma-Aldrich. The following antibodies were used for immunoblotting and/or immunofluorescence assays: rabbit anti-FLAG M2 (Cat. No. F7425, Sigma-Aldrich), rabbit anti-EXO5 (Cat. No. LS-C167355, LifeSpan BioSciences), rabbit anti-phospho-histone H2A.X (Ser139) (Cat. No. ab2893, Abcam), rabbit anti-H2A.X (Cat. No. 2595S, Cell Signaling), rabbit anti-histone H3 (Cat. No. 4499, Cell Signaling), rabbit anti-GFP (Cat. No. NB600-308, Novus Biologicals), mouse anti-PSA (Cat. No. SC-7316, Santa Cruz Biotechnology), rabbit anti-AR (Cat. No. Cell Signaling), rabbit anti-GAPDH (Cat. No. 2118S, Cell Signaling), anti-RPA32/RPA2 antibody [9H8] (Cat. No. ab2175, Abcam), secondary goat anti-rabbit HRP (Cat. No. SC-2004, Santa Cruz Biotechnology), and secondary goat anti-mouse HRP (Cat. No. GTX213111-01, GeneTex, Inc.).

PCa sample collection and whole-exome sequencing analysis

Blood samples from 20 families with metastatic PCa were collected by Dr. Theodore Krontiris, M.D., Ph.D. and Jeffrey N. Weitzel, M.D. Each family ($n = 20$) comprised three or seven brothers affected by metastatic PCa. All families studied were Caucasian. All procedures involving human subjects were approved by the City of Hope Institutional Review Board. Written informed consent was obtained from all participants. Genomic DNA was isolated from peripheral blood using the DNeasy Blood and Tissue kit (Qiagen Inc). Whole-exome sequencing was performed using an Illumina HiSeq 2500 (Illumina, Inc.). Reads were aligned to the hg19 genome assembly using the Burrow-Wheeler Aligner (BWA). Variants were identified using the Genome Analysis Toolkit (GATK) pipeline and annotated using the AnnoVar software package. A mean coverage depth of 99.7x per sample was achieved, with 86.7% of targets covered at a depth of ≥20, and revealed around 100,000 single nucleotide variations and insertion/deletion mutations in each sample.

***In silico* prediction of the functional impacts of variants using bioinformatics tools**

The pathogenicities of missense mutations were predicted using six online bioinformatics tools. The SIFT method was used to predict deleterious effects based on the degree of conservation of amino acid residues in sequence alignments derived from closely related sequences and collected through PSI-BLAST [1]. A SIFT score < 0.05 suggests that a given substitution is deleterious. PolyPhen-2 was used to predict the possible effects of an amino acid substitution on the structure and function of a human protein using physical and comparative considerations [2]. The LRT identified conserved amino acid positions and deleterious mutations using a comparative genomics dataset of multiple vertebrate species [3]. MutationTaster was used to evaluate the disease-causing potential of DNA sequence alterations using a Naive Bayes classifier that integrates information, such as evolutionary conservation, splice-site changes, loss of protein features, and changes that might affect the amount of mRNA, from various biomedical databases [4]. MutationAssessor was used to predict the functional impacts of amino acid substitutions in

proteins based on evolutionary conservation of the affected amino acid in protein homologs [5]. The CADD method was used to score the deleteriousness of a base substitution in the human genome [6]. CADD score quantitatively prioritize functional, deleterious, and disease causal variants across a wide range of functional categories, effect sizes and genetic architectures. A mutation with a CADD score >20 (top 1% of relative deleterious variants) is considered deleterious. Mutations predicted to be pathogenic or deleterious using three or more bioinformatics tools were considered potentially deleterious.

Analysis of data from the Database of Genotypes and Phenotypes (dbGaP)

The genotyping data of 2507 participants in a distribution set (phs001391.v1.p1.c8) was requested from a central data repository (dbGaP) established by the National Institutes of Health (NIH). This dataset covers a total of 505219 genome-wide SNPs in samples from 1591 PCa cases and 916 healthy controls of European, Asian, and African ancestry. The contributing institutions established that the data can be distributed, categorized as "Disease-Specific (Prostate Cancer)" (DS-PC). Standard GWAS data quality control procedures for genetic markers and subjects were performed using PLINK v1.07 (pngu.mgh.harvard.edu/~purcell/plink). Quality control procedures excludes the samples and SNPs based on the following criteria: SNP call rate < 95%, Hardy-Weinberg equilibrium test $P < 1 \times 10^{-6}$ in the control population, frequency filtering (MAF < 0%), and participant call rate < 95%.

Sequence alignment and 3D structure prediction

EXO5 protein sequences were downloaded from the NCBI database. Sequence alignment of eight EXO5 orthologs from different species and Cas4 from *Sulfolobus solfataricus* was performed using MEGA X (v10.0.4) [7]. The 3D homology modeled structure of the EXO5 N-terminus was downloaded from the online SWISS-MODEL Repository [8]. The homology modeled structure of the EXO5 C-terminal (amino acids 334–365) was built using the SWISS-MODEL online server [9] with mouse DNA2 as a template (Protein Data Bank ID: 5EAN) [10]. Pymol software [11] was used to align the models using the combinatorial extension (CE) algorithm and to generate the structural figures.

Protein expression, purification, and Western blot analysis

The human wild-type and mutant EXO5 proteins were purified using a mammalian expression system. The human *EXO5* gene (Cat. No. HG15535, Sino Biological) was cloned into a 3xFlag-tagged mammalian expression vector. Site-directed mutagenesis, performed using the QuickChange mutagenesis kit, was used to introduce the L151P mutation. Introduction of the mutation was confirmed by direct sequencing. Wild-type and mutant EXO5 proteins were expressed in HEK293T cells and proteins purified using anti-FLAG antibodies attached to M2 magnetic beads using the standard protocol (Cat. No. M8823, Sigma-Aldrich). Protein purity was checked by SDS-PAGE followed by Coomassie blue staining.

Cells were trypsinized and centrifuged. Cell pellets were resuspended in 2.5x volume of RIPA buffer (20 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 0.5% sodium deoxycholate, and 0.5% SDS containing protease and phosphatase inhibitors). Total protein concentrations in the lysates were determined using the Protein Assay Kit (Bio-Rad) and measuring absorption at 595 nm on the microplate reader (Bio-Rad). 40–60 µg of each protein sample was separated in an SDS polyacrylamide gel (10–15%). The protein samples were transferred (semidry transfer) from the gel onto a nitrocellulose membrane (Bio-Rad) and blocked for 1 h in blocking solution (5% milk powder in PBS with 0.1% Tween [PBST]). The membrane was incubated with primary antibodies overnight at 4^{o}C, washed with PBST three times, incubated with secondary antibodies for 1 h, and finally washed with PBST. The blots were visualized using an ECL assay (Thermo Scientific).

Nuclease activity assays

Nuclease activity assays were performed in 10-µl reactions containing 40 mM Tris-HCL (pH 7.6), 25 mM KCl, 5 mM $MgCl_2$ (PCR grade), 2 mM DTT, 2% glycerol, 100 µg/ml BSA, and 0.5 pmol of ^{32}P-end-labeled single-flap DNA substrates (Supplementary Table 3). The reaction mixture was incubated at 37^{o}C for multiple time points (0, 5, 10, 20, 40, 60, or 80 min). The reactions were stopped by adding 2x reaction stop buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol), followed by heat denaturation (5 min, 95^{o}C). The reaction mixture was resolved on 7M urea-15% polyacrylamide gels and the bands were visualized using autoradiography.

Immunofluorescence assays

Immunofluorescence staining for γH2AX was performed on parental and nuclease-defective LNCaP cells, grown to 80% confluency on coverslips in 12-well plates. For androgen deprivation, cells were grown for 48–72 h in charcoal-stripped serum and phenol red-free RPMI medium. After deprivation, cells were treated with 100 nM DHT for 2 h with or without the TOP2 inhibitor VP-16 (10 µM, 30 min), as described previously [12]. After up to 24 h of recovery in media without DHT, cells were fixed with 4% paraformaldehyde, permeabilized with 0.3% Triton-X100, and blocked using blocking buffer (1X PBS, 3% BSA, 0.3% Triton X-100). Image-iT FX signal enhancer was used to block background staining that resulted from non-specific interactions of the fluorescent dye with the cells. All antibodies were diluted in a dilution buffer (1X PBS, 1% BSA, 0.3% Triton X-100). The cells were incubated with anti-phospho-histone γH2AX (Ser139) antibodies at 4^{o}C for 1 h. Slides were washed three times with PBS and goat anti-rabbit antibodies labeled with Alexa Fluor 555 (Cat. No. A21428, Thermo Fisher Scientific). Nuclei were stained with DAPI (Cat. No. D1306, Thermo Fisher Scientific). Coverslips were mounted onto slides using SlowFade Gold antifade reagent (Cat. No. P36930, Thermo Fisher Scientific) and examined using a Zeiss Observer wide-field fluorescence microscope.

Cellular I-SceI/GFP reporter-based HDR and NHEJ assays

The HDR and NHEJ reporters (DR-GFP and EJ5-GFP, respectively) [13] were integrated into the LNCaP cell line. Single clones of the reporter cell lines were used to generate knockout lines. 1.0 x 10^5 cells of parental and nuclease-defective LNCaP cell lines integrated with either the HDR (LNCaP-HDR) or NHEJ (LNCaP-NHEJ) cassettes were grown in 12-well plates for 24 h. Cells were transfected with the I-SceI expression vector (pCBASce, 1.0 µg) using Lipofectamine 2000 (2.5 µl) in OptiMEM media (1 ml). The transfection media was removed after 6–8 h and replaced with complete RPMI 1640 media containing 10% fetal bovine serum and 1% penicillin-streptomycin. After 48–72 h of incubation at 37ºC and 5% CO_2, cells were trypsinized and fixed using 10% buffered formalin. The frequency of GFP-positive cells was determined by flow cytometry using the BD LSRFortessa (BD Biosciences). The fold decrease in repair efficiency was calculated by dividing the percentage of GFP-positive nuclease-defective cells by the percentage of GFP-positive parental cells. The results are presented as mean ± standard deviation of at least three independent transfections.

End resection assay

The end resection assay, which measures RPA protein bound to single-stranded DNA, was performed as described previously [14, 15]. Briefly, parental or *EXO5* knockout LNCaP cells were seeded in 6-well plates and treated with or without the cross-linking agent CPT (1 µM, 1 h) and VP-16 (100 µM, 1 h). Cells were collected by trypsinization and washed with PBS. Cells were permeabilized using 0.2% TritonX-100 in 1x PBS on ice for 7 min, washed with 0.1% BSA-PBS, and fixed with 4% fresh formaldehyde for 15 min at room temperature. Cells were further incubated with anti-RPA32/RPA2 antibodies (1:200) for 1 h on ice, washed with 0.1% BSA-PBS, and resuspended in anti-mouse secondary antibodies (1:200) labeled with Alexa Fluor 555 for 30 min. Cells were finally washed and resuspended in DAPI solution containing 0.02% sodium Azide, 250 ng/ml RNase A, and 2 µg/ml DAPI for 30 min at 37ºC. Staining of cells was analyzed using BD LSRFortessa (BD Bioscience). For androgen deprivation, cells were grown for 48–72 h in charcoal-stripped serum and phenol red-free RPMI medium. Cells were then treated with 100 nM DHT for 2 h with or without V-16 (100 µM, 1 h). For overexpression of wild-

type *EXO5*, 2 µg of a vector containing the *EXO5* gene was transfected into cells during seeding. In all experiments, cells were treated three days after transfection.

Reverse transcription PCR to assess TMPRSS2-ERG gene fusion

Gene fusion assay was performed using a modified protocol as described by Tomlins [16]. Briefly, parental and nuclease defective LNCaP cells were seeded as $1x10^6$ cells in 60 mm plates. Cells were androgen deprived for 48-72 h by culture in charcoal-stripped serum and phenol red-free RPMI medium. Cells were treated with 100nM DHT with or without irradiation (1Gy, IR) and allowed to grown at 37ºC and 5% CO_2 for 24 h [12]. Total RNA was isolated using Trizol, 2 µg of RNA was used as a template for reverse transcription to obtain cDNA using M-MLV reverse transcriptase (Promega) and primers specific to exon 11 of the *ERG* gene (Supplementary Table 3). For RT-PCR, 100 ng of cDNA was PCR amplified using JumpStart Taq Ready Mix (Sigma-Aldrich) in a 25 µl total reaction volume using primers specific to exon 1 of *TMPRSS2* and exon 6 of *ERG*. Products were resolved on 1.5% agarose gels using electrophoresis. Gene fusion products were confirmed by Sanger sequencing.

Cell proliferation and cellular migration assay

For cell proliferation assay, $1.5x10^5$ cells of parental and Exo5 knockout LNCaP cell lines were seeded in duplicates in 10cm plate in complete media. Cells were incubated at 37ºC for 3, 5 and 7 days and the media was refreshed every 3-4 days. Cells proliferation was measured by trypsinization followed by trypan blue vital staining using hemocytometer. For cell migration assay, $1.5x10^5$ parental and Exo5 knockout LNCaP cells were seeded per well in 6-well culture plate. After 48 h, a wound was created with a 100ul pipette tip and cells were incubated in complete medium. Wound healing process was monitored visually by taking picture at day 0, 3, 5 using EVOS FL Imaging System (Life Technologies). For overexpression of wild-type *EXO5*, vector containing the *EXO5* gene was transfected into cells a day before and next day cells were trypsinized and seeded into wells/plates. Experiments were repeated at least two times.

Statistics analysis

The results obtained from at least three biological replicates are presented as mean with error bars representing the standard deviation. Statistical analysis was performed using unpaired, two-tailed t-tests to evaluate differences between two groups. A *P*-value < 0.05 was considered significant and marked with an asterisk (*) in the relevant figures.

References

1 Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research* 2003; 31: 3812-3814.

2 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P *et al*. A method and server for predicting damaging missense mutations. *Nat Methods* 2010; 7: 248-249.

3 Chun S, Fay JC. Identification of deleterious mutations within three human genomes. *Genome Res* 2009; 19: 1553-1561.

4 Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010; 7: 575-576.

5 Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic acids research* 2011; 39: e118.

6 Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics* 2014; 46: 310-315.

7 Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular biology and evolution* 2018; 35: 1547-1549.

8 Bienert S, Waterhouse A, de Beer TA, Tauriello G, Studer G, Bordoli L *et al*. The SWISS-MODEL Repository-new features and functionality. *Nucleic acids research* 2017; 45: D313-D319.

9 Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R *et al*. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* 2018; 46: W296-W303.

10 Zhou C, Pourmal S, Pavletich NP. Dna2 nuclease-helicase structure, mechanism and regulation by Rpa. *eLife* 2015; 4.

11 DeLano WL. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography* 2002; 40: 82-92.

12 Haffner MC, Aryee MJ, Toubaji A, Esopi DM, Albadine R, Gurel B *et al*. Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. *Nature genetics* 2010; 42: 668-675.

13 Gunn A, Stark JM. I-SceI-based assays to examine distinct repair outcomes of mammalian chromosomal double strand breaks. *Methods Mol Biol* 2012; 920: 379-391.

14 Forment JV, Walker RV, Jackson SP. A high-throughput, flow cytometry-based method to quantify DNA-end resection in mammalian cells. *Cytometry Part A : the journal of the International Society for Analytical Cytology* 2012; 81: 922-928.

15 Howard SM, Yanez DA, Stark JM. DNA damage response factors from diverse pathways, including DNA crosslink repair, mediate alternative end joining. *PLoS genetics* 2015; 11: e1004943.

16 Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW *et al*. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 2005; 310: 644-648.

Supplementary Table 1: Sequences of primers used for Sanger sequencing.

Primer name	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
ATM	TTGAATGAATGTTGTTTCTAGGTCC	GTCAGATAGCTGGTTGTTGGC
ATM_del	CTGTAACTCCAGGTGGTTCC	GACAGCTGTCAGCTTTAA
BRCA2	GATGGCAGTGATTCAAGTAAAAATG	GTCACTAGTTGATTTCCAGTACCA
BRIP1	AGTTGAATCAGCATACTCAAGTGAA	CTTGCCGTAGTCACATTGGC
DCLRE1B	AGGGTGACTATGTTGACGGC	ATTCCATAGGCACGAGGACC
EXD2	TGCCAGCTGTATGACCTTTG	AGGTTGCAGTGAGCCAAGAC
ERCC2	CCCCGAATGACCTTCTGTCC	CCGTATAATCCAGAGCGGGC
EXO1	GGTTGACACAGATGTAGCACG	ATAGCTCGTCATTCACATGTAGG
EXO5	CTGGACCTAGAAGATGCCCAAG	TTAGGCTAGCAGGGGTCACTT
EXOG	AGTTAGTTACCAGGTGATTGGCG	GCCTGCTTCATTACAGACGG
FANCB	CCCATGACTCCAGATTTCGG	ATAGCCGGGTGATGTGATTTGA
FANCF	CAGCGCTGTCTCCTGTCTAT	AAATCCGTGACACTACATGCCA
FANCM	GCCATGTTCAGCTTGGTAATGTG	AGTAGAACGCACCTTGGACTG
HFM1	TTCTACAGCCATAGAGTAGAGAGG	ACTCAGAAAAAGTGAGTAGCTGAAC
MMS19	TCTTTAGCAAGGAGGCGTTGG	CTGATTCAGTTGGACGCAGG
MSH6	ATGGGGGAGATCGTTGGAC	AAAGGGCCTCATGACCTGAATG
MUS81	AGCTCAGTTAGTTACGGGGT	AAGACAGGGTAGGCTCAGTG
MUTYH	TGTGAGCAAAGGTGGAGCC	GATTCTCAGGGAATGGGGGC
PAPOLA	AGTTAAAATCCTTGGGGCAGGA	CTCTCTTAAGTATCCCTCTTCTCC
PARP2	ACTGACGTCATGAGGAAAGCC	CCATACAACCTGTCACCCCA
POLD1	GTGACCCACATCTTAGCCCC	TCTAGGATCTGGGGTGACGG
POLH	GTGCCTGGCCAAAAATGTGT	ACCCGGCCTATGTGTAATAATTT
POLI	TGCTTGACGTCTTGCACATC	ACCTTGGCCTAGATGTTTTGGA
PRIMPOL	GCCAGGAGTTCAAGATCAGC	CCAGACAGCAAAATTGCAAAC
RAD23A	CAGAAGCCAGGGTCCGATT	CTAATGCACAGACCTCCTGCT
RAD51AP1	AAGCATGGATTTGCCTACACAGAA	CAACTAGAAGCCACGGGTAGT
RAD51C	CTTGGCTCACTGCAACCTCT	GTCTTGAACTCCTGGCCTCA
REV3L	GCATCATACCACAGACTCAGC	CCTGGGCTTTTCTGGTACATCA
RFC1	CACAGAAGACATGTGCCTTGAAA	GACTTTGGCCTGCCAGATAAAAA
SLX4	TCAGCGAAAGCTTCTCCAGG	GGACGACCCACTTGTGTGAT
XPC	AGGTTAGCTGACATTTAAGATCTGG	GCTCTTACCGGTCTGAGTTG
XRCC5	GGAGGTCCAGAGCAGAAGATAC	AGCCCAGAAGGTAGCTCTAAATC

Supplementary Table 2: Details of mutations of DNA damage response and repair pathway genes identified in PCa families.

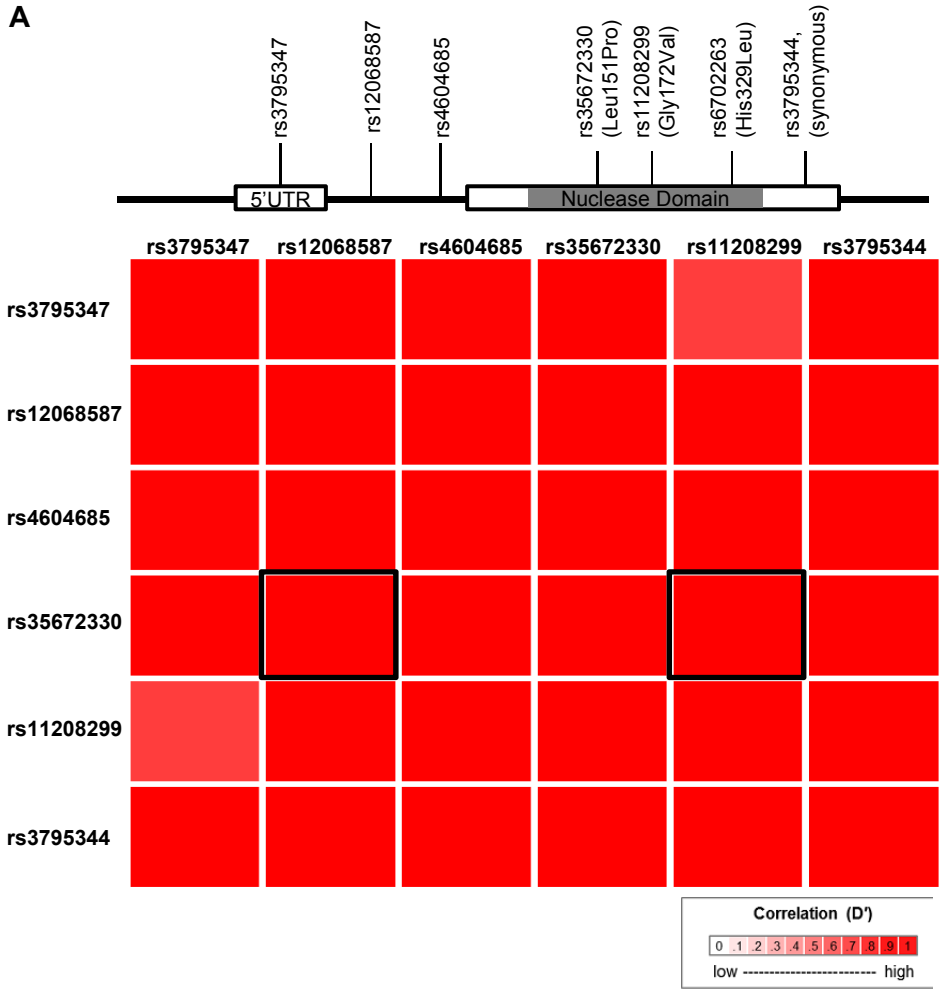
Gene	Chr:	Position[a]	Base pair change	rs ID[b]	Amino acid substitution	Gl-AF[c]	Eu-AF[d]
ATM	11	108153549	A>G	rs587782195	NM_000051.3 p.Asn1230Ser	N/A	NA
ATM	11	108214064	ATTTCAG TGCC>A	rs71855130	NM_000051.3:p.Asp2795fs/ c.8385_8394delTTTCAGTGCC	N/A	N/A
BRCA2	13	32912750	G>T	rs28897727	NM_000059.3 p.Asp1420Tyr	0.39	1
BRIP1	17	59924512	G>A	rs4988346	NM_032043.2 p.Val193Ile	0.13	0
DCLRE1B	1	114454742	A>T	rs35397235	NM_022836.3 p.Asn510Tyr	0.02	0
ERCC2	19	45855468	G>A	N/A	NM_000400.3 p.Arg730Gln	N/A	N/A
EXD2	14	69697203	G>C	N/A	NM_001193360.1 p.Cys202Ser	0.04	N/A
EXO1	1	242042437	G>A	rs4149978	NM_006027.4 p.Arg634Gln	0.14	0
EXO5	1	40980668	T>C	rs35672330	NM_022774.1 p.Leu151Pro	1.96	6
EXOG	3	38565576	G>T	rs1141223	NM_005107.3 p.Gly277Val	0.46	1
FANCB	X	14863136	T>C	rs142959373	NM_001018113.1 p.Phe590Ser	0.05	0
FANCF	11	22646398	C>T	rs45451294	NM_022725.3 p.Pro320Leu	0.49	1
FANCM	14	45605463	A>G	rs61746895	NM_020937.2 p.Thr77Ala	1.62	1
FANCM	14	45606290	C>T	rs77374493	NM_020937.2 p.Thr176Ile	0.14	1
FANCM	14	45636328	A>G	rs61753893	NM_020937.2 p.Asn655Ser	0.66	N/A
FANCM	14	45645715	A>G	rs45604036	NM_020937.2 p.Asn1253Ser	1.30	3
FANCM	14	45658024	C>T	rs61746943	NM_020937.2 p.Thr1600Ile	0.62	3
FANCM	14	45665661	A>G	rs45557033	NM_020937.2 p.Asn1876Ser	0.86	3
HFM1	1	91816318	C>G	N/A	NM_001017975.4 p.Ala728Gly	0.02	N/A
MMS19	10	99225645	C>T	rs12360068	NM_001289405.1 p.Ala558Val	1.57	5
MSH6	2	48030685	C>T	rs63750442	NM_000179.2 p.Thr1100Met	N/A	0
MUS81	11	65629482	G>A	rs148465534	NM_025128.4 p.Arg139Gln	0.18	1
MUTYH	1	45795084	C>T	rs140118273	NM_001128425.1 p.Ser515Phe	0.46	1
PAPOLA	14	96986546	G>C	N/A	NM_032632.4 p.Glu55Gln	N/A	N/A
PARP2	14	20822308	A>G	rs3093921	NM_005484.3 p.Asp235Gly	0.73	3
POLD1	19	50919693	C>T	rs374016016	NM_001256849.1 p.Thr954Met	N/A	N/A
POLH	6	43565568	G>T	rs2307456	NM_006502.2 p.Gly209Val	0.42	1
POLI	18	51807260	A>G	rs3218784	NM_007195.2 p.Ile261Met	0.79	2
PRIMPOL	4	185582954	T>C	rs142122035	NM_152683.3 p.Val102Ala	0.44	2
RAD23A	19	13059626	C>T	rs4987202	NM_005053.3 p.Thr200Met	0.26	0
RAD51AP1	12	4657293	A>G	rs61731949	NM_001130862.1 p.Met136Val	0.18	0
RAD51C	17	56798128	A>G	rs28363317	NM_058216.2 p.Thr287Ala	N/A	1
REV3L	6	111694124	G>C	rs3218599	NM_002912.4 p.Asp1812His	0.91	3
RFC1	4	39343994	A>G	N/A	NM_001204747.1 p.Gln101Arg	N/A	N/A
SLX4	16	3658880	G>A	rs149117119	NM_032444.2 p.Arg29His	0.81	0
XPC	3	14214524	C>T	rs2229089	NM_004628.4 p.Leu48Phe	0.99	2
XRCC5	2	217012977	G>A	rs35408277	NM_001290268.1 c.349-55G>A	0.10	0

[a]Human genome assembly GRCh37.p13/hg19
[b]dbSNP141 rs ID
[c]1000 Genome project alternative allele frequency in the global population
[d]1000 Genome project alternative allele frequency in European populations

Supplementary Table 3: Sequences and details of guide RNA for CRISPR-Cas9, primers for RT-PCR, and DNA substrates for biological assays.

Name	Direction	Sequence (5'-3')
***EXO5* guide RNA for CRISPR Cas9**		
GuideSeq-1	Forward	caccgTTCTCTAGCTAGGTGTATGC
GuideSeq-1	Reverse	aaacGCATACACCTAGCTAGAGAAc
GuideSeq-2	Forward	caccgATTCCTACCCTGCAGTCAGA
GuideSeq-2	Reverse	aaacTCTGACTGCAGGGTAGGAATc
***EXO1* guide RNA for CRISPR Cas9**		
GuideSeq-1	Forward	caccGCTTCAGAACCCATCCATGTG
GuideSeq-1	Reverse	aaacCACATGGATGGGTTCTGAAGC
GuideSeq-2	Forward	caccGTAGCTGTGGATACATATTGC
GuideSeq-2	Reverse	aaacGCAATATGTATCCACAGCTAC
RT-PCR primers for *TMPRSS2-ERG*		
TMPRSS2-F	Forward	GAGCTAAGCAGGAGGCGGA
ERG-R1	Reverse	TATTCTTTCACCGCCCACTC
ERG-R2 (Exon 11)	Reverse	ACTGCCAAAGCTGGATCTGG
***Beta-actin* primers**		
Actin-b	Forward	ACCATGGATGATGATATCGCCG
Actin-b	Reverse	CACCATCACGCCCTGGTG
***PSA* primers**		
PSA-F	Forward	CACACCCGCTCTACGATATG
PSA-R	Reverse	CAGGTCCATGACCTTCACAG
***SF3A3* primers**		
SF3A3	Forward	TACGAAAGGAGGAGCTCAATGCAA
SF3A3	Reverse	AGATCTCATTTGGGTGCTTCCGGT
DNA substrates for biochemical assays		
CBRF8-20		
GAAGTGCCATTAATACCGGTAGTTAGGACTGCTTGACATCCCAAGCAGACCTATCTTAAC		
BRF1		
GTTAAGATAGGTCTGCTTGGGATGTCAAGCAGTCCTAACTGGAAATCTAGCTCTGTGGAGT TGAGGCAGAGTCCTTAAGC		
CBRF8		
CAGTTGTTGAATGCAAAGAAGAAGTGCCATTAATACCGGTAGTTAGGACTGCTTGACATC CCAAGCAGACCTATCTTAAC		
CBRF120		GCTTAAGGACTCTGCCTCAA

A

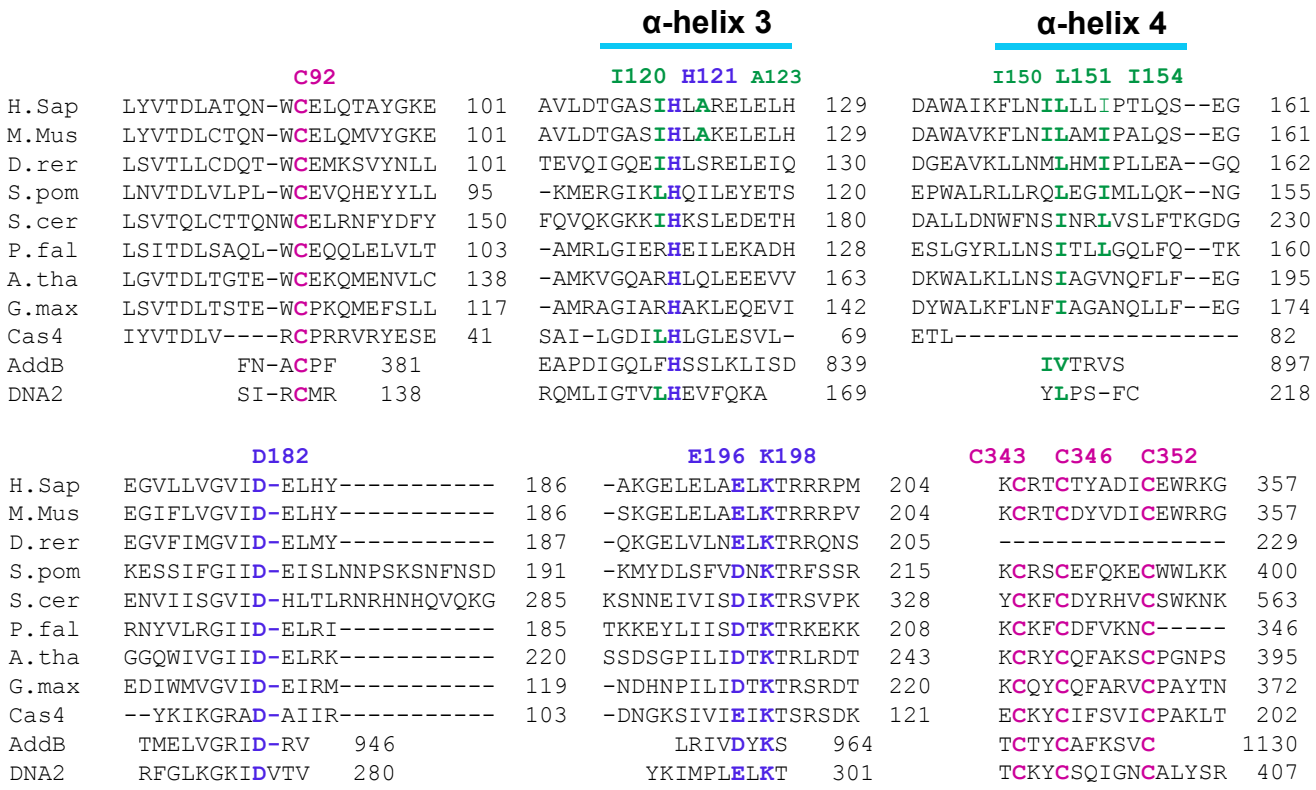


B

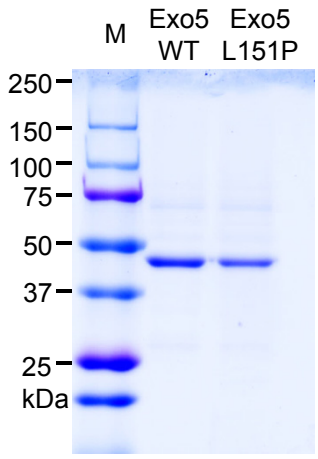
D' values

RS_number	rs3795347	rs12068587	rs4604685	rs35672330	rs11208299	rs3795344
rs3795347	1	1	0.966	1	1	1
rs12068587	1	1	1	1	1	1
rs4604685	0.966	1	1	1	1	1
rs35672330	1	1	1	1	1	1
rs11208299	1	1	1	1	1	1
rs3795344	1	1	1	1	1	1

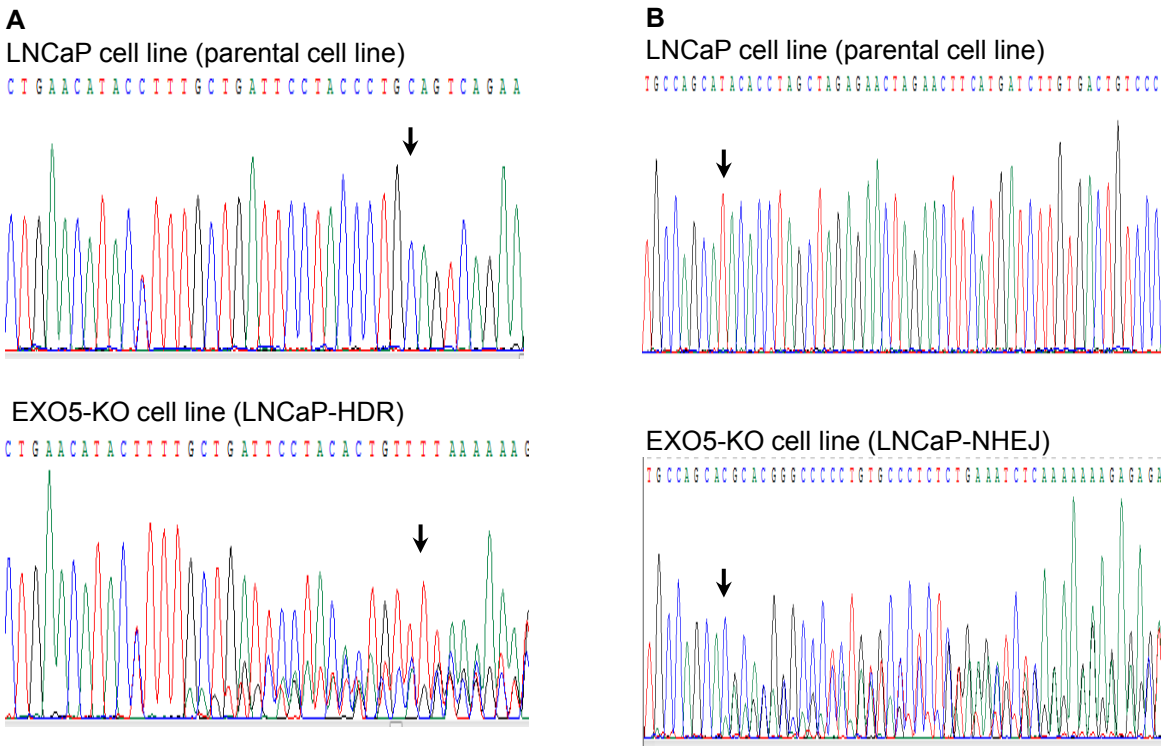
Supplementary Figure 1. Pairwise linkage disequilibrium analysis of the *EXO5* L151P mutation and SNPs that are associated with PCa susceptibility in the Database of Genotypes and Phenotypes (dbGaP). (A) Linkage disequilibrium structure (based on D' value) for HapMap data from a population of European ancestry (CEU) for the *EXO5* L151P mutation (rs35672330) and other *EXO5* SNPs. Data was downloaded from NCBI (https://ldlink.nci.nih.gov/). (B) Pairwise D' values between the *EXO5* L151P mutation and other SNPs.



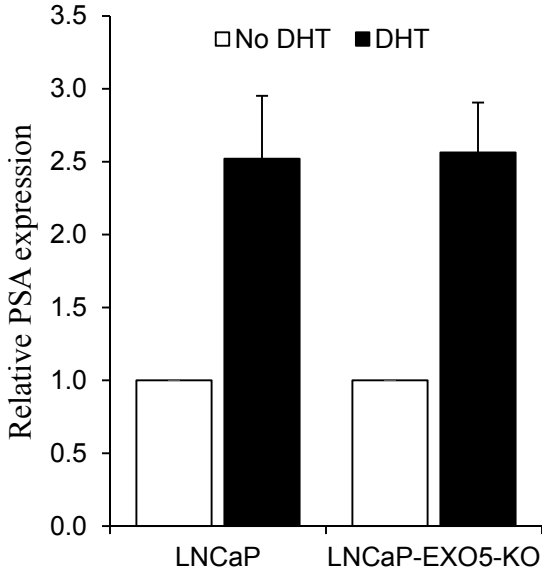
Supplementary Figure 2. EXO5 amino acid sequence alignment across species. Sequence alignment of EXO5 in eight species [H. sap (*Homo sapiens*), M. mus (*Mus musculus*), D. rer (*Danio rerio*), S. pom (*Schizosaccharomyces pombe*), S. cer (*Saccharomyces cerevisiae*), P. Fal (*Plasmodium falciparum*), A. tha (Arabidopsis thaliana), and G. max (*Glycine max*)]. Cas4 (*Sulfolobus Solfataricus*), AddB (*Bacillus subtilis*), and human DNA2 genes are also included based on partial sequence alignment and 3D structure homology. EXO5 L151 and four cysteine residues are highly conserved among the species. Three residues, histidine at 121, aspartic acid at 182, and glutamic acid at 196, that bind to Mg^{++} are also relatively conserved. The DNA2 sequence was aligned to observe cysteine residue conservation.



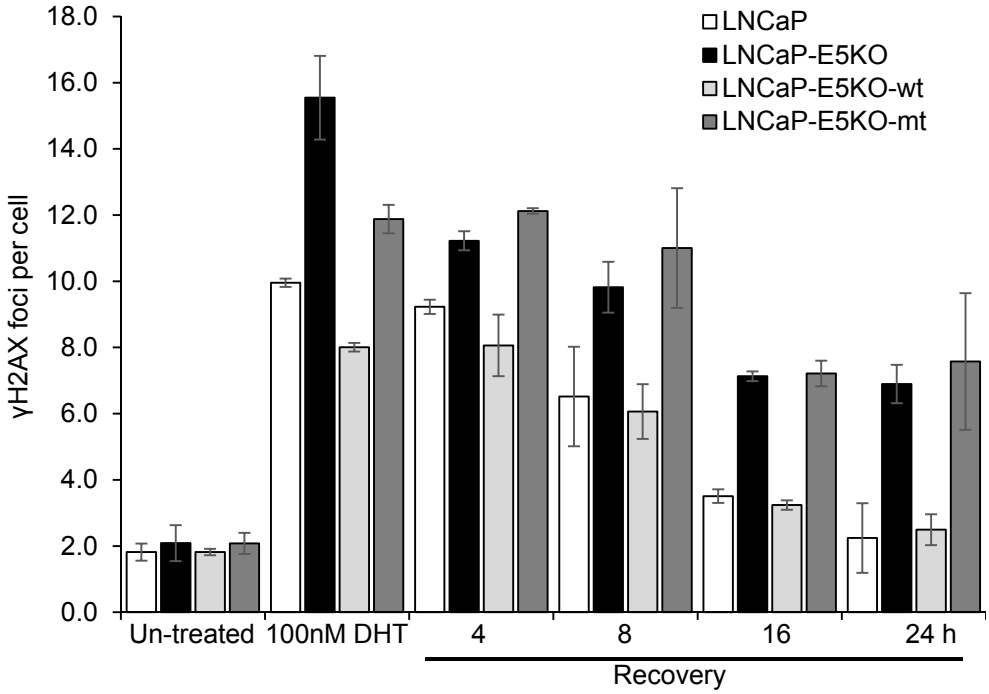
Supplementary Figure 3. EXO5 protein expression and purification. Commassie blue-stained gel of purified FLAG-tagged wild-type (WT) and L151P mutant EXO5. M: protein markers in kDa.



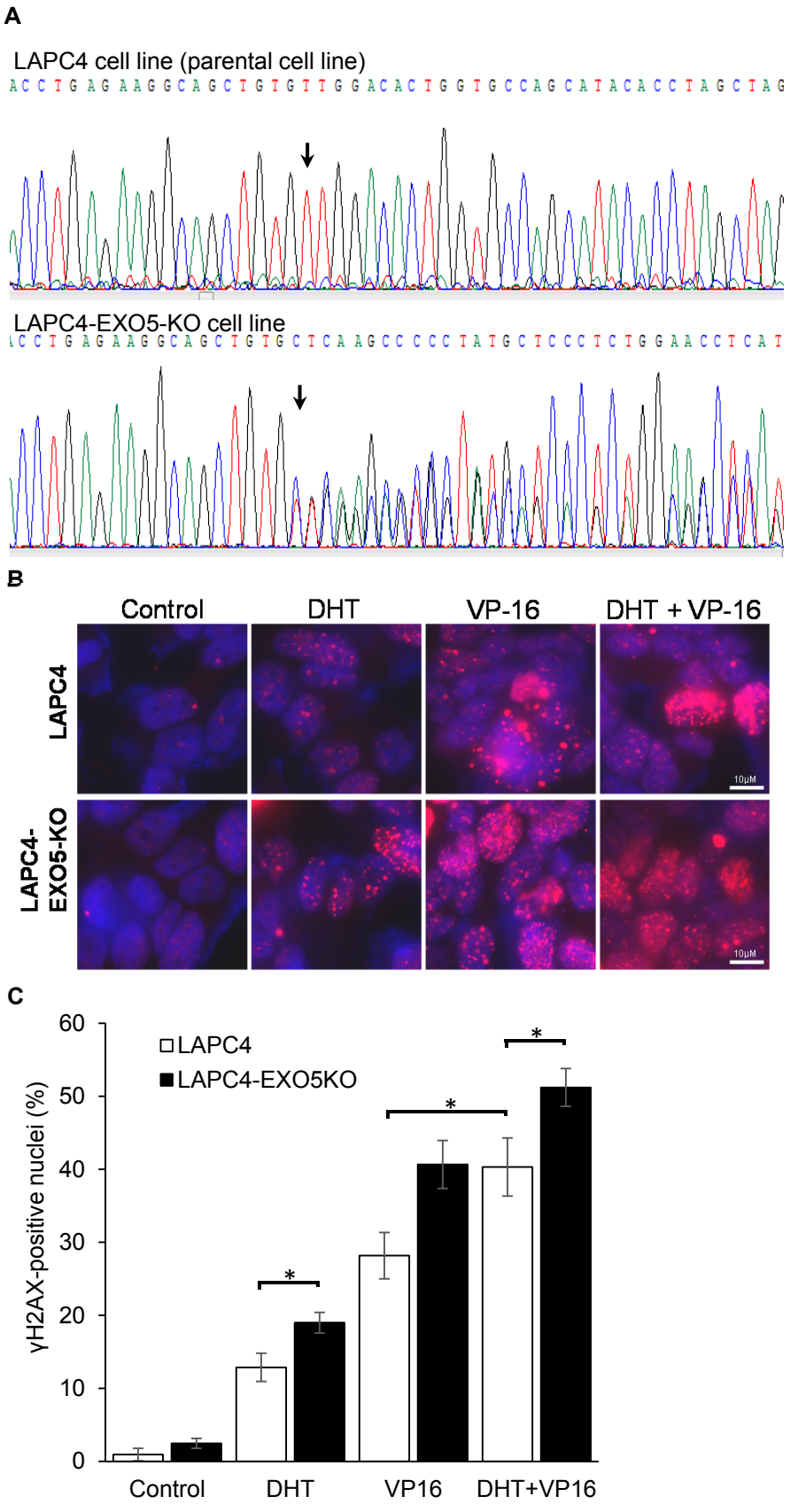
Supplementary Figure 4. Sequence confirmation of *EXO5* knockout cell lines. (A) Sequence validation of *EXO5* knockout in LNCaP cells containing HDR cassettes. Coding region of EXO5 gene was disrupted using specific guideRNA in CRISPR-Cas9 technique. The *EXO5* gene was PCR amplified using primers specific to the gene (Supplementary Table 3). The PCR product was purified and sequenced using Sanger sequencing. Arrows indicate the position of sequence disruption in both strands of the coding region of the gene. (B) Sequence validation of *EXO5* knockout in LNCaP cells containing NHEJ cassettes. Coding region of EXO5 gene was disrupted using specific guideRNA in CRISPR-Cas9 technique. The *EXO5* gene was PCR amplified using primers specific to the gene (Supplementary Table 3). The PCR product was purified and sequenced using Sanger sequencing. Arrows indicate the position of sequence disruption in both strands of the coding region of the gene.



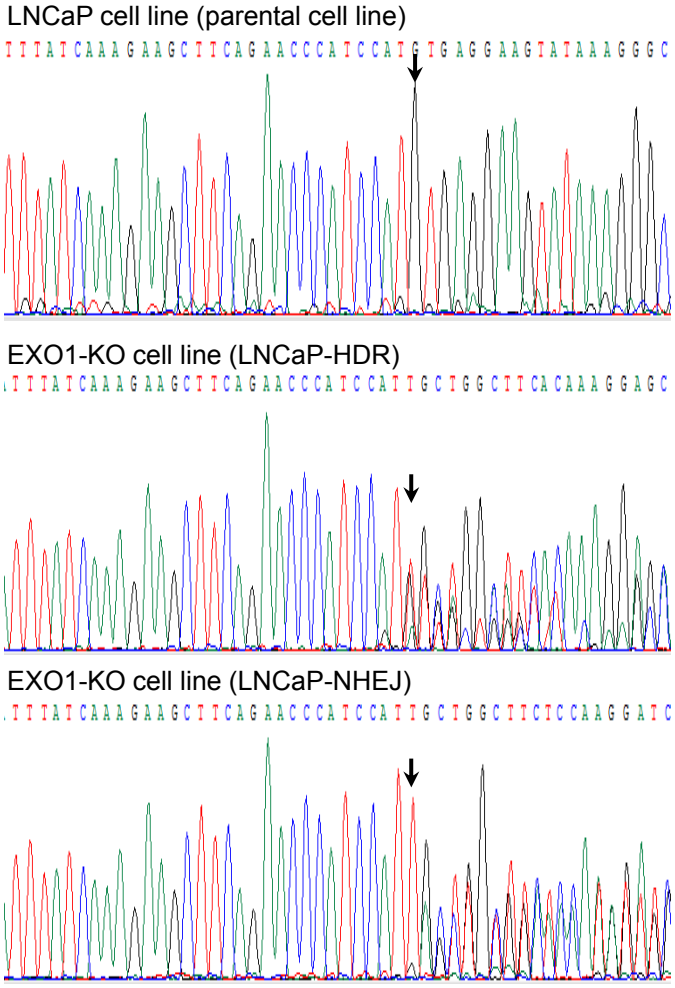
Supplementary Figure 5. DHT stimulation to LNCaP cells. Quantitative PCR analysis (Fold expression change) of prostate specific antigen (PSA) in parental (WT) and EXO5 knockout (EX5KO) LNCaP cells with and without DHT (100 nM, 2 h) stimulation. Expression values of PSA was normalized to SF3A3 gene as internal control. Experiment was repeated at least two times.



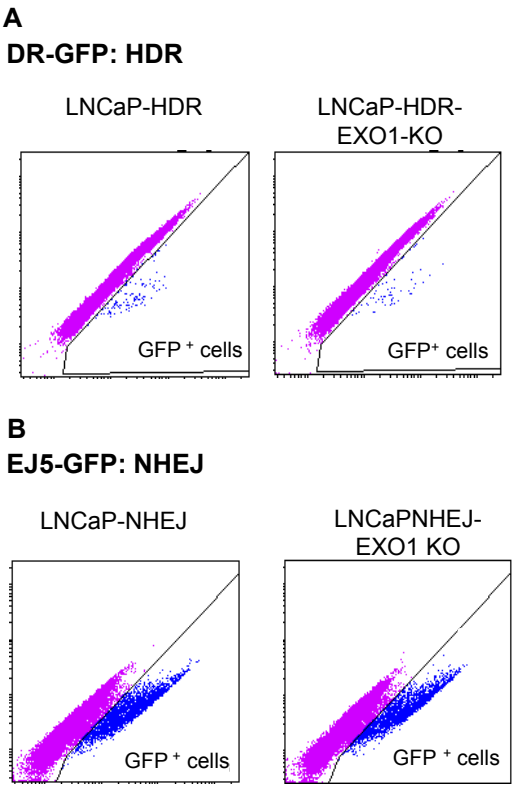
Supplementary Figure 6. *EXO5* knockout LNCaP cells show decrease in androgen induced DSB repair capacity. Time-course assay of γH2AX foci resolution using immunoflourscence assay from LNCaP and LNCaP-EXO5KO cells, overexpressing WT or MT EXO5, untreated and after DHT treatment (100 nM, 2 h) or recovery (incubation in media without DHT) for the indicated times. The graph shows quantification of average γH2AX foci per cell and at least 100 cells were analyzed. Data shown are mean ± standard deviation of 3 independent experiments. **P*< 0.05.


A
LAPC4 cell line (parental cell line)
LAPC4-EXO5-KO cell line
B
Control
DHT
VP-16
DHT + VP-16
LAPC4
LAPC4-EXO5-KO
10μM
C
LAPC4
LAPC4-EXO5KO
γH2AX-positive nuclei (%)
60
50
40
30
20
10
0
Control
DHT
VP16
DHT+VP16
*

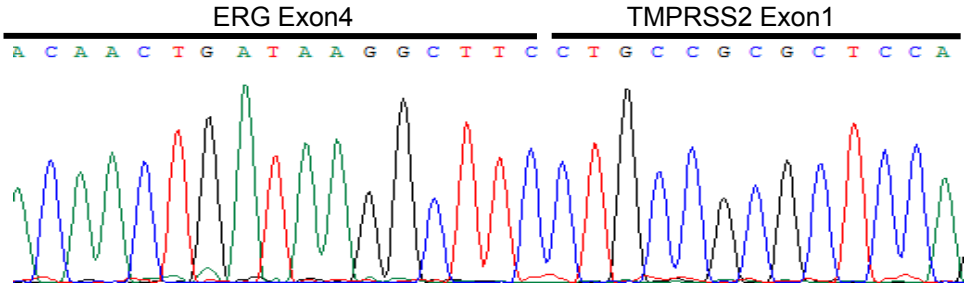
Supplementary Figure 7. Knockout of EXO5 promotes DHT-induced DNA damage in androgen-sensitive LAPC4 PCa cells. (A) Sequence validation of *EXO5* knockout in LAPC4 cells. Coding region of EXO5 gene was disrupted using specific guideRNA in CRISPR-Cas9 technique. The *EXO5* gene was PCR amplified using primers specific to the gene (Supplementary Table S3). The PCR product was purified and sequenced using Sanger sequencing. Arrows indicate the position of sequence disruption in both strands of the coding region of the gene. (B) Immunofluorescence images of parental LAPC4 and EXO5-KO after treatment with DHT (100 nM, 2 h) and/or the TOP2 inhibitor VP-16 (10μM, 30 min), stained for γH2AX. (C) The quantification of γH2AX-positive nuclei (10 or more foci present in nucleus) per field in LAPC4 and EXO5-KO cells treated as described in Figure 3F. Data shown are mean ± standard deviation of three fields per experiments from 3 independent experiments. *$P < 0.05$.



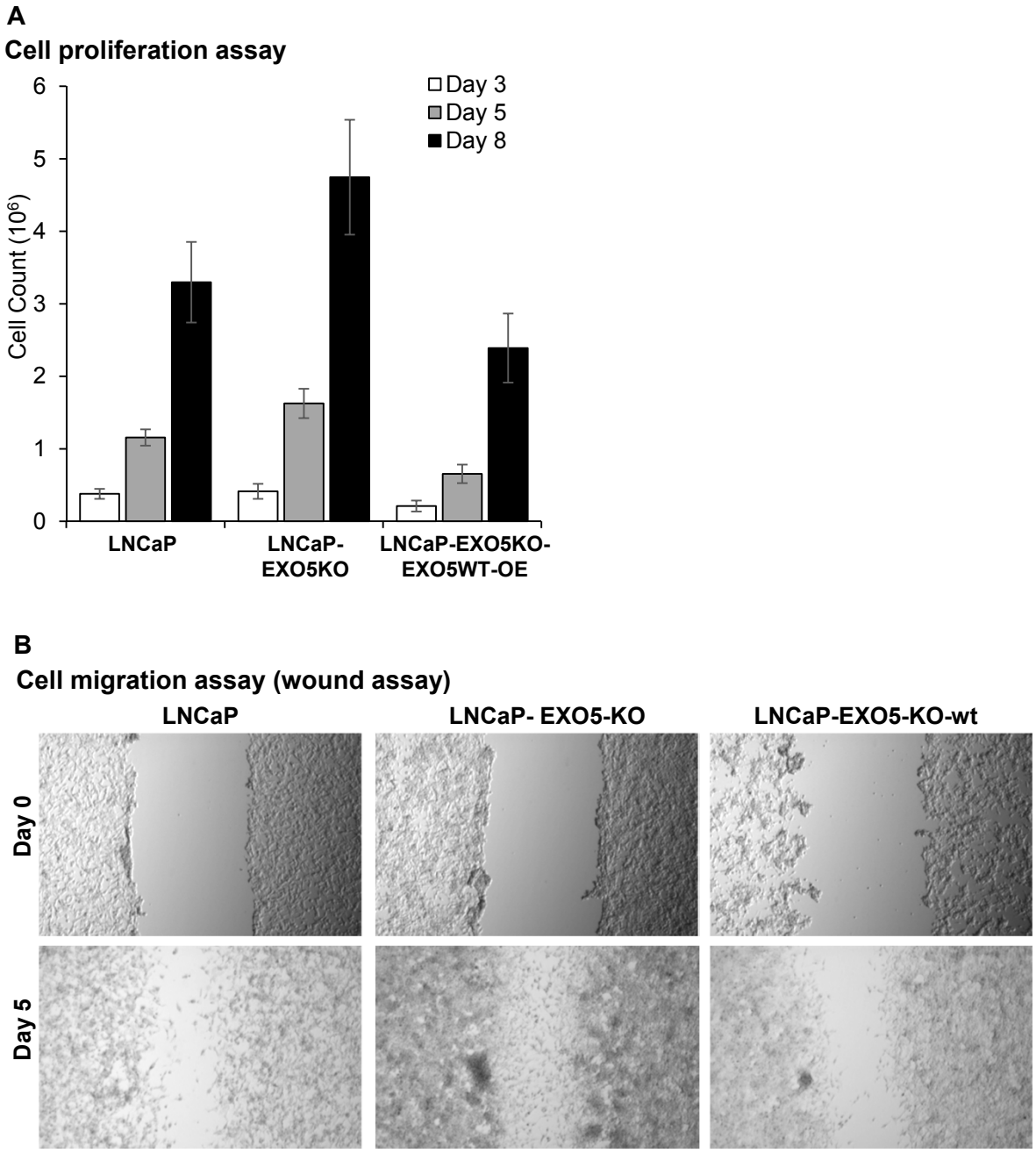
Supplementary Figure 8. Sequence confirmation of *EXO1* knockout cell lines. Sequence validation of *EXO1* knockout in parental LNCaP cells and LNCaP cells containing HDR or NHEJ cassettes. Coding region of EXO1 gene was disrupted using specific guideRNA in CRISPR-Cas9 technique. The *EXO1* gene was PCR amplified using primers specific to the gene (Supplementary Table 3). The PCR product was purified and sequenced using Sanger sequencing. Arrows indicate the position of sequence disruption in both strands of the coding region of the gene.



Supplementary Figure 9. Representative flow cytometry profiles to estimate DNA repair efficiency in *EXO1* knockout cells. (A) Representative flow cytometry profile of untransfected and I-SceI-transfected LNCaP cells containing HDR GFP reporter cassettes (LNCaP-HDR). Cells showing green fluorescence greater than autofluorescence were gated to determine the percentage of GFP^+ cells. Knockout of the *EXO1* gene in LNCaP-HDR cells significantly reduced HDR efficiency. (B) Representative flow cytometry profile of untransfected and I-SceI-transfected LNCaP cells containing NHEJ GFP reporter cassettes (LNCaP-NHEJ). Cells showing green fluorescence greater than autofluorescence were gated to determine the percentage of GFP^+ cells. Knockout of the *EXO1* gene in LNCaP-NHEJ cells did not reduce NHEJ repair efficiency.



Supplementary Figure 10. TMPRSS2-ERG fusion gene sequencing. Sequence confirmation of the *TMPRSS2-ERG* fusion gene using automated Sanger DNA sequencing.



Supplementary Figure 11. Knockout of EXO5 in prostate cell lines promotes cell proliferation and cellular migration. (A) $1.5x10^5$ cells of each group were plated in 10cm plate in duplicates and counted at day 3, 5 and 8. (B) Representative image demonstrate cell migration of parental, Exo5 knockout and wild type Exo5 complemented LNCaP cells at day0 and day5.