

# Supporting Information

He et al. 10.1073/pnas.1421415111

## SI Materials and Methods

**Cell Culture.** Human cells lines were obtained from American Type Culture Collection (ATCC) via the Tissue and Cell Culture Core Laboratory at Baylor College of Medicine, where they are regularly submitted for cell line authentication by short tandem repeats profiling and mycoplasma testing, and were passaged for less than 6 mo. HEK293 T cells, cervical carcinoma HeLa cells, and PC DU145 cells were cultured in DMEM high glucose (Life Technologies) with 10% FBS (Life Technologies). LNCaP and 22Rv1 PC cells and the human immortalized prostatic cell line PNT1a (Sigma-Aldrich) were cultured in Roswell Park Memorial Institute medium 1640 (RPMI 1640; Life Technologies) supplemented with 10% FBS. LAPC4 cells were cultured in Iscove's Modified Dulbecco's Medium (Life Technologies) plus 15% FBS, 1 nM R1881, and 2 mM of L-glutamine. PC3 cells were cultured in DMEM/F12 (F-12 Nutrient Medium; Life Technologies) with 10% FBS. VCaP cells were maintained in DMEM high glucose (Life Technologies) with 10% FBS and 1 nM R1881. The LNCaP-derived CRPC cell line Abl [characterized by and obtained from Zoran Culig (Medical University of Innsbruck, Innsbruck, Austria)] (1) was maintained in phenol red-free RPMI medium 1640 (Life Technologies) supplemented with 10% charcoal-stripped FBS. The LNCaP-derived CRPC cell line C4-2 [characterized by and provided by Leland W. K. Chung (Cedars-Sinai Medical Center, Los Angeles)] (2) was grown in T-Medium (Life Technologies) supplemented with 5% FBS. RWPE-1 epithelial cells derived from the peripheral zone of a histologically normal adult human prostate and immortalized via human papilloma virus 18 were cultured in Keratinocyte Serum-Free Medium according to the instructions of the ATCC. LNCaP cells with inducible expression of AR-FL or AR3/v7 have been described previously (3). All media were supplemented with 100 U/mL penicillin and 100 µg/mL streptomycin. Cells were maintained in a 5% CO<sub>2</sub> incubator at 37 °C.

**Characterization of the AR Transcriptional Complex *ex Vivo*.** A 400-bp ARE-harboring sequence from the human *KLK3/PSA* gene promoter was PCR-amplified using biotin-labeled primers, purified, and incubated with 150 ng of SRC-2 and p300 proteins and 1 µM R1881 in the absence or presence of 200 ng AR protein and 500 ng GATA2 protein on ice for 1 h in buffer containing 20 mM Hepes, 150 mM KCl, 0.1% Nonidet P-40, 8% glycerol, 1 mM DTT, and protease inhibitor mixture. The reaction mixture then was incubated with 10 µL of Dynabeads M-280 Streptavidin beads (Life Technologies) for 15 min at RT. After three washes using reaction buffer, 10 µL of 2× sample buffer was added to the Dynabeads, followed by boiling and immunoblot analysis.

**IHC.** PC tissue microarrays (TMAs) have been described previously (4–6). IHC was performed using a previously described mouse monoclonal anti-GATA2 antibody (7). Antigen retrieval was performed in Tris-EDTA buffer, pH 8.0, for 20 min in a steamer, followed by cooling at RT for 20 min. Endogenous peroxidase was blocked using hydrogen peroxide in methanol. Nonspecific staining was blocked using Rodent Block M (Biocare Medical) for 1 h at RT. Slides were incubated with primary antibody at 1:800 dilution overnight at 4 °C. GATA2 was detected using MM HRP Polymer (Biocare Medical) for 30 min at RT followed by Stable DAB Plus (Diagnostic BioSystems) for 5 min at RT. Slides then were counterstained using hematoxylin (Biocare Medical) for 15 s. All rinses were made using Tris-

Tween Buffer, pH 7.4. The immunostained slides were reviewed by a pathologist (D.G.R.), who followed the TMA map to record a score for each sample. The reviewer was blinded to all sample identifiers. The equipment consisted of a Nikon Eclipse 80i microscope connected to a CCD digital camera (Nikon digital sight ds-fi1) and ImageJ image analysis software [National Institutes of Health (NIH), [imagej.nih.gov/ij/](http://imagej.nih.gov/ij/)]. Counting criteria and software settings were identical for all slides within the same assay. The average of the results from replicate core samples from each tumor specimen was considered for each case. Cores containing more than 20% of tumor were included for quantitation. Cores with less than 20% of tumor and missing cores were not included in the final analysis. The immunostain was represented as total cytoplasmic stained area (brown color) and was expressed in pixels, and total integrated optic density was expressed in arbitrary optic density units. For statistical analysis, all cases displaying similar total integrated optic density (mean ± SE) were grouped together in a four-tier scale (0–3+) intensity score. Negative staining was defined as absence of cytoplasmic stain (blue color) and was given a score of 0. Percent of positive tumor area was calculated as the ratio of positive stain (brown stain) to total stain (brown and blue stain). Labeling frequency was scored as 0 (0%), 1 (1–33%), 2 (34–66%), or 3 (67–100%). Tumor index was calculated by multiplying the intensity score by the extent score for each case to yield an intensity index (0–9).

## Gene-Expression Profiling after Silencing GATA2 and AR in PC Cells.

To determine the transcriptomic footprint of GATA2 and AR in PC cells, LNCaP cells were transfected with Stealth Select pre-designed siRNA (Life Technologies) for *GATA2*, *AR*, or non-target (NT) control, respectively, for 72 h at a final concentration of 50 nM in complete medium (supplemented with 10% FBS), using Lipofectamine RNAiMAX (Life Technologies) according to the manufacturer's instructions. Total RNA was extracted using TRIzol (Life Technologies), purified with the RNeasy Mini Kit (Qiagen) following the manufacturer's instructions, and analyzed on an Affymetrix Human Exon 1.0 ST Array platform at the Genomic and RNA Profiling Core at Baylor College of Medicine. Gene-expression data were normalized by using the Bioconductor *lumi* (8) package using the R statistical system. Gene-expression differences were inferred using the *t* test and imposing a fold change exceeding 4/3× ( $P < 0.05$ ), using the R statistical system.

## Prognostic Significance of GATA2 Activity in PC Patient Cohorts.

We applied our GATA2 gene signature (as derived above by comparing samples treated with si-GATA2 with samples treated with si-NT) to previously reported expression datasets from two large PC patient cohorts for which clinical outcomes have been collected: Taylor et al. (9), and TCGA. For each gene in the GATA2 signature and for each PC specimen, we computed the *z*-score for its expression within each cohort, as described previously (9), and computed the sum *z*-score for each specimen. Specifically, the *z*-scores of genes repressed by GATA2 (i.e., up-regulated by si-GATA2) were subtracted from the *z*-scores of genes induced by GATA2 (i.e., down-regulated by si-GATA2), resulting in a corresponding GATA2 activity score for each specimen. Within each dataset, specimens were ranked according to their GATA2 activity score, and association with biochemical-free recurrence survival was evaluated by comparing the top 25% of the ranked specimens with the bottom 75% of the ranked

specimens using the log-rank test. Survival significance was assessed by using the package *survival* (10) in the R statistical system.

**GSEA.** GSEA was carried out using the GSEA software package (11) to assess the degree of similarity between the studied gene signatures. For the si-GATA2 experiment, genes were ranked by the fold change between the si-GATA2 and si-NT samples. We also examined the GATA2 signature against the REACTOME Pathways collection maintained at the MSigDB ([www.broadinstitute.org/gsea/msigdb/](http://www.broadinstitute.org/gsea/msigdb/)); we used adjusted  $q < 0.05$  as our filtering criterion. For the gene signature derived from the AR siRNA, we used both the down-regulated genes (positively regulated by AR) and up-regulated genes (suppressed by AR). The NES and adjusted  $q$ -values were computed using the GSEA method, based on 1,000 random permutations of the ranked genes. We also downloaded a previously reported transcriptomic signature of AR3/v7 in LNCaP cells (12) from the Gene Expression Omnibus (GEO) (GSE36549). We ranked the genes by the fold-change between AR3/v7-expressing and control LNCaP cells and performed GSEA analysis vs. our GATA2 signature.

**Correlation of GATA2 Activity with AR Activity in PC Patient Cohorts.** We calculated the GATA2 activity and AR activity scores (derived from the si-GATA2 and si-AR gene signatures, respectively, as described above) for human PC specimens using gene-expression datasets from the previously reported cohorts Taylor et al. (9) and Cai et al. (13). For each specimen, we plotted its GATA2 activity and AR activity scores on the  $x$  and  $y$  axis, respectively, and computed the Pearson correlation coefficient  $r$  and associated  $P$  value using the R statistical system.

**ChIP-Seq and Data Analysis.** LNCaP cells grown to 70–80% confluency in complete medium (supplemented with 10% FBS) were fixed in 1% formaldehyde for 10 min. ChIP, post-ChIP quality control by PCR, library generation, postamplification quality control by PCR, and sequencing by Illumina Genome Analyzer were performed as previously described (14, 15). The following antibodies were used for ChIP: rabbit polyclonal to AR (39782; Active Motif); goat polyclonal against FOXA1 (ab5089; Abcam); rabbit polyclonal against SRC-1 (sc-8995; Santa Cruz); rabbit polyclonal against SRC-2 (GRIP-1, sc-8996; Santa Cruz); goat polyclonal against SRC-3 (sc-7216; Santa Cruz); rabbit polyclonal against c-myc (sc-764; Santa Cruz); rabbit polyclonal against RNA pol II phosphoserine 2 (Ab5095; Abcam); rabbit polyclonal against p300 (sc-585; Santa Cruz); and rabbit polyclonal against CBP (sc-369; Santa Cruz). We also used two antibodies against GATA2: a rabbit polyclonal antibody GATA2-r (sc9008; Santa Cruz) and a mouse monoclonal antibody GATA2-m (provided by T.M.), both of which previously had been used successfully for ChIP-Seq in cancer cell lines (7, 16). Unprecipitated LNCaP chromatin served as input control. Each of the LNCaP GATA2, FOXA1, AR, SRC1, SRC2, SRC3, p300, CBP, cMyc, pol II, and input ChIP-Seq samples yielded 31–40 million raw reads 36 nt long; the reads were mapped onto the human genome University of California Santa Cruz (UCSC) hg19/National Center for Biotechnology Information build 37 using the Burrows–Wheeler aligner algorithm (17), with 23–33 million reads mapping uniquely. High-resolution genomewide maps were derived and visualized in the UCSC Genome Browser and in the Integrative Genomic Viewer (18, 19). Finally, the macs2 (20) peak-calling package was used to identify enriched regions using default parameters; LNCaP input reads were used as negative control. Motif analysis was performed using MEME-ChIP (21) and the HOMER tools (22) with default parameters. To determine the enrichment of a feature set with respect to our ChIP-Seq peak calls, we used a permutation testing across the human genome; we generated random sets with the same number of peaks and the same peak size by using a uniform random dis-

tribution across the human genome, determining the amount of peak overlap expected by chance and the SD. We used in-house scripts for this analysis. We also integrated additional PC ChIP-Seq datasets for p300, MED12, H3K4me1, H3K4me2, and H3K27ac from Wang et al. (23), for histone H3 acetylation, H3K4me3, H3K9me3, H3K27me3, and H3K36me3 from Yu et al. (24), for RNA pol II and ERG from Asangani et al. (25), as well as DNase I hypersensitivity-Seq data from He et al. (26) and WDR5 and H3T11-p from GSE55279 (27). We also downloaded from GEO (GSE58478) ChIP-Seq data for ARVs and AR-FL generated in 22Rv1 cells by Lu et al. (28) and compared them with our GATA2 cistrome (generated in LNCaP cells).

Because we used two different antibodies against GATA2 (GATA2-r and GATA2-m) for ChIP-Seq, we compared the two resulting ChIP-Seq datasets. We first computed sequence tag coverage over 1-kb genome tiling windows using in-house scripts and then assessed the Pearson correlation coefficient between the coverage for each antibody using the R statistical system. As described in *Results* in the main text, the two GATA2 antibodies gave very strong correlation. Subsequent bioinformatic comparisons were performed using the dataset obtained with the GATA2-r antibody (sc9008; Santa Cruz Biotechnologies).

To quantify the distribution of ChIP-Seq peaks of GATA2, FOXA1, and AR with respect to gene features, we used the UCSC RefSeq gene model. We inferred enhancers in LNCaP cells treated with androgen as described previously in Wang et al. (23). We determined peaks' overlap with gene features or enhancers using the BEDTOOLS software (29).

**ChIP and RT qPCR.** To validate the in vivo recruitment of GATA2 and AR to the chromatin at select gene loci, ChIP assays were performed using the ChIP-IT Express Kit (Active Motif) with minor modifications. Briefly, LNCaP cells were androgen-starved for 48 h, followed by treatment with R1881 or vehicle (ethanol) as indicated, and then were cross-linked with 1% formaldehyde (final concentration) and lysed to release chromatin. The chromatin was sonicated, quantified, and was incubated with GATA-2 (GATA2-r and GATA2-m) antibody or AR antibody or normal rabbit IgG (nonspecific antibody control) overnight at 4 °C. The immune complexes were precipitated with protein-A Dynabeads (Life Technologies), followed by extensive washing as recommended by the manufacturer. The chromatin–protein–antibody complexes were eluted, and the DNA–protein cross-links were reversed; then the chromatin DNA pulled down by the antibody was purified with the Qiaquick PCR purification kit (Qiagen). The specific protein-binding genomic DNA sequences of the genes of interest were detected by RT qPCR using the Power SYBR Green PCR Master Mix reagents (Life Technologies) on a StepOne Plus Real-time PCR System (Life Technologies). The sequences of the qPCR primer sets used in our experiments are shown in Table S2. The abundance of the detected DNA (relative concentration) was calculated and normalized to each of its total input (before immunoprecipitation) amounts, respectively. The normalized relative DNA concentration in each sample was expressed as the fold change over its respective control. Each experiment was repeated at least three times, and the results were analyzed for statistical significance using the paired-sample  $t$  test. The differences between samples with or without treatment were considered significant if the  $P$  value was less than 0.05.

**FRET Imaging.** CFP/YFP FRET experiments were performed as described previously (30, 31) with slight modifications. ECFP-AR-FL, ECFP-AR-V7 (kindly provided by Marco Marcelli, Baylor College of Medicine, Houston, TX), and EYFP-GATA2 were transiently expressed in HeLa cells cultured on standard 12-mm glass coverslips. Twenty-four hours after transfection, the cells were treated with 2 nM DHT or ethanol for an additional 18 h. The cells were fixed in 4% paraformaldehyde (30 min),

quenched in 100 mM NH<sub>4</sub>Cl (10 min), and mounted with SlowFade Gold (Life Technologies). FRET imaging was performed with the Leica TCS SP5 confocal microscope with a 63×/1.4 oil objective lens. Single-plane images were imaged with a frame size of 512 × 512 pixels. After imaging, FRET calculations were performed using PixFRET (32). FRET measurements on individual nuclei were acquired on single-focal planes of the derived FRET image. Spectral bleedthrough was corrected for by acquiring specimens expressing each CFP or YFP construct individually. Standard values for  $\alpha$  and  $\beta$  coefficients were 0.81 (CFP) and 0.12 (YFP) acquired from single-donor/acceptor plasmid expression experiments.

**MTT Assay.** To quantify the impact of GATA2 siRNA and GATA SMI on PC cell growth/viability, we used the MTT assay, as previously described (33).

**K7174 and Proteasome Inhibitor Treatment.** LNCaP cells were plated in six-well plates (0.5 million cells per well) and allowed to attach overnight. Then cells were treated with 250 nM of bortezomib and/or 20  $\mu$ M of K7174 (synthesized by Baylor College of Medicine, Houston, TX) for 4 h. At the end of treatment, cells were harvested, and total cell lysates were prepared. Immunoblot analyses were performed as indicated.

**SDS/PAGE and Immunoblot Analyses.** Cell lysates from siRNA- or K7174-treated cells were separated on polyacrylamide gels; then the proteins were transferred to PVDF membranes and detected by immunoblotting using monoclonal antibodies or specific anti-sera, as indicated. Rabbit polyclonal anti-GATA2 and AR antibodies were obtained from Cell Signaling. Monoclonal anti-AR3/v7-specific antibody was obtained from Precision Antibody. Anti- $\beta$ -actin antibody was obtained from Sigma-Aldrich. Blots were washed with 1× PBS and Tween-20 and were incubated with anti-mouse or anti-rabbit HRP-conjugated secondary antibodies for 1 h. Protein signals were detected with SuperSignal Western chemiluminescent substrate (Thermo Scientific, Inc.) and developed with X-ray films according to the manufacturer's instructions (Thermo Scientific, Inc.) followed by densitometric analysis using NIH ImageJ image analysis software (34) or on a

Molecular Imager VersaDoc MP 4000 System (Bio-Rad) using QuantiOne software (Bio-Rad) for quantitative analysis of the band intensity.

**RT qPCR.** Total RNA was isolated from cultured cells using TRIzol (Life Technologies) and purified using an RNeasy Mini Kit (Qiagen) following the manufacturer's instructions. Total RNA was quantified and reverse transcribed with a High Capacity Reverse Transcription kit (Life Technologies). The resulting cDNA was combined with sensiFAST SYBR No ROX kit (Bio-line) or SYBR Green PCR Master mix (Life Technologies) and primer pairs to detect specific transcripts by qPCR. For RT qPCR analysis of 18S and AR transcripts, a TaqMan One-Step RT-PCR Master Mix Kit (Life Technologies) was used. Taqman assays for eukaryotic 18S rRNA (endogenous control) and for AR mRNA quantification (hs0090244) were obtained from Life Technologies. The RT qPCRs were performed on a StepOne Plus Realtime PCR System (Life Technologies), and amplification data were processed as previously described (33). Values were normalized to total RNA or 18S rRNA or  $\beta$ -actin, as indicated. Each experiment was repeated at least three times, and the results were analyzed for statistical significance.

Differences in transcript levels between samples were analyzed using *t* test and considered significant if the *P* value <0.05.

**Recombinant Protein Production for DNA Pull Downs.** The recombinant AR, SRC2, and P300 proteins used in these studies were produced by the Baylor College of Medicine Baculovirus Core Facility using baculoviral vectors expressed in SF9 insect cells. Recombinant AR was produced in the presence of 10 nM of R1881, and cell lysates were purified using a FLAG affinity column. Recombinant SRC2 containing an N-terminal His tag and a C-terminal FLAG tag was purified from cell lysates using a nickel-NTA column and a FLAG affinity column. Recombinant His tagged-p300 was produced and cell lysates were purified using a nickel-NTA column following the manufacturer's protocol. Recombinant GATA2-11R was purchased from LD BioPharma, Inc.

- Culich Z, et al. (1999) Switch from antagonist to agonist of the androgen receptor bicyclutamide is associated with prostate tumour progression in a new model system. *Br J Cancer* 81(2):242–251.
- Wu TT, et al. (1998) Establishing human prostate cancer cell xenografts in bone: Induction of osteoblastic reaction by prostate-specific antigen-producing tumors in athymic and SCID/bg mice using LNCaP and lineage-derived metastatic sublines. *Int J Cancer* 77(6):887–894.
- Krause WC, Shafi AA, Nakka M, Weigel NL (2014) Androgen receptor and its splice variant, AR-V7, differentially regulate FOXA1 sensitive genes in LNCaP prostate cancer cells. *Int J Biochem Cell Biol* 54:49–59.
- Agoulnik IU, et al. (2005) Role of SRC-1 in the promotion of prostate cancer cell growth and tumor progression. *Cancer Res* 65(17):7959–7967.
- Agoulnik IU, et al. (2006) Androgens modulate expression of transcription intermediary factor 2, an androgen receptor coactivator whose expression level correlates with early biochemical recurrence in prostate cancer. *Cancer Res* 66(21):10594–10602.
- Ayala G, et al. (2004) High levels of phosphorylated form of Akt-1 in prostate cancer and non-neoplastic prostate tissues are strong predictors of biochemical recurrence. *Clin Cancer Res* 10(19):6572–6578.
- Kanki Y, et al. (2011) Epigenetically coordinated GATA2 binding is necessary for endothelium-specific endomucin expression. *EMBO J* 30(13):2582–2595.
- Du P, Kibbe WA, Lin SM (2008) lumi: A pipeline for processing Illumina microarray. *Bioinformatics* 24(13):1547–1548.
- Taylor BS, et al. (2010) Integrative genomic profiling of human prostate cancer. *Cancer Cell* 18(1):11–22.
- Therneau TM, Grambsch PM (2000) *Modeling Survival Data: Extending the Cox Model* (Springer, New York), pp xiii, 350 pp.
- Subramanian A, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102(43):15545–15550.
- Hu R, et al. (2012) Distinct transcriptional programs mediated by the ligand-dependent full-length androgen receptor and its splice variants in castration-resistant prostate cancer. *Cancer Res* 72(14):3457–3462.
- Cai C, et al. (2013) ERG induces androgen receptor-mediated regulation of SOX9 in prostate cancer. *J Clin Invest* 123(3):1109–1122.
- Labhart P, et al. (2005) Identification of target genes in breast cancer cells directly regulated by the SRC-3/AIB1 coactivator. *Proc Natl Acad Sci USA* 102(5):1339–1344.
- Stashi E, et al. (2014) SRC-2 is an essential coactivator for orchestrating metabolism and circadian rhythm. *Cell Reports* 6(4):633–645.
- Tijssen MR, et al. (2011) Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev Cell* 20(5):597–609.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178–192.
- Robinson JT, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24–26.
- Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9):R137.1–137.9.
- Machanic P, Bailey TL (2011) MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics* 27(12):1696–1697.
- Heinz S, et al. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38(4):576–589.
- Wang D, et al. (2011) Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 474(7351):390–394.
- Yu J, et al. (2010) An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* 17(5):443–454.
- Asangani IA, et al. (2014) Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature* 510(7504):278–282.
- He HH, et al. (2012) Differential DNase I hypersensitivity reveals factor-dependent chromatin dynamics. *Genome Res* 22(6):1015–1025.
- Kim JY, et al. (2014) A role for WDR5 in integrating threonine 11 phosphorylation to lysine 4 methylation on histone H3 during androgen signaling and in prostate cancer. *Mol Cell* 54(4):613–625.

28. Lu J, et al. (2014) The cistrome and gene signature of androgen receptor splice variants in castration-resistant prostate cancer cells. *J Urol* 50022-5347(14):04214-1.
29. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841-842.
30. Hartig SM, He B, Long W, Buehrer BM, Mancini MA (2011) Homeostatic levels of SRC-2 and SRC-3 promote early human adipogenesis. *J Cell Biol* 192(1):55-67.
31. Hartig SM, et al. (2012) Feed-forward inhibition of androgen receptor activity by glucocorticoid action in human adipocytes. *Chem Biol* 19(9):1126-1141.
32. Feige JN, Sage D, Wahli W, Desvergne B, Gelman L (2005) PixFRET, an ImageJ plug-in for FRET calculation that can accommodate variations in spectral bleed-throughs. *Microsc Res Tech* 68(1):51-58.
33. Geng C, et al. (2013) Prostate cancer-associated mutations in speckle-type POZ protein (SPOP) regulate steroid receptor coactivator 3 protein turnover. *Proc Natl Acad Sci USA* 110(17):6997-7002.
34. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671-675.
35. Wang XD, et al. (2007) Identification of candidate predictive and surrogate molecular markers for dasatinib in prostate cancer: Rationale for patient selection and efficacy monitoring. *Genome Biol* 8(11):R255.1-255.11.
36. Creighton MP, et al. (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* 107(50):21931-21936.

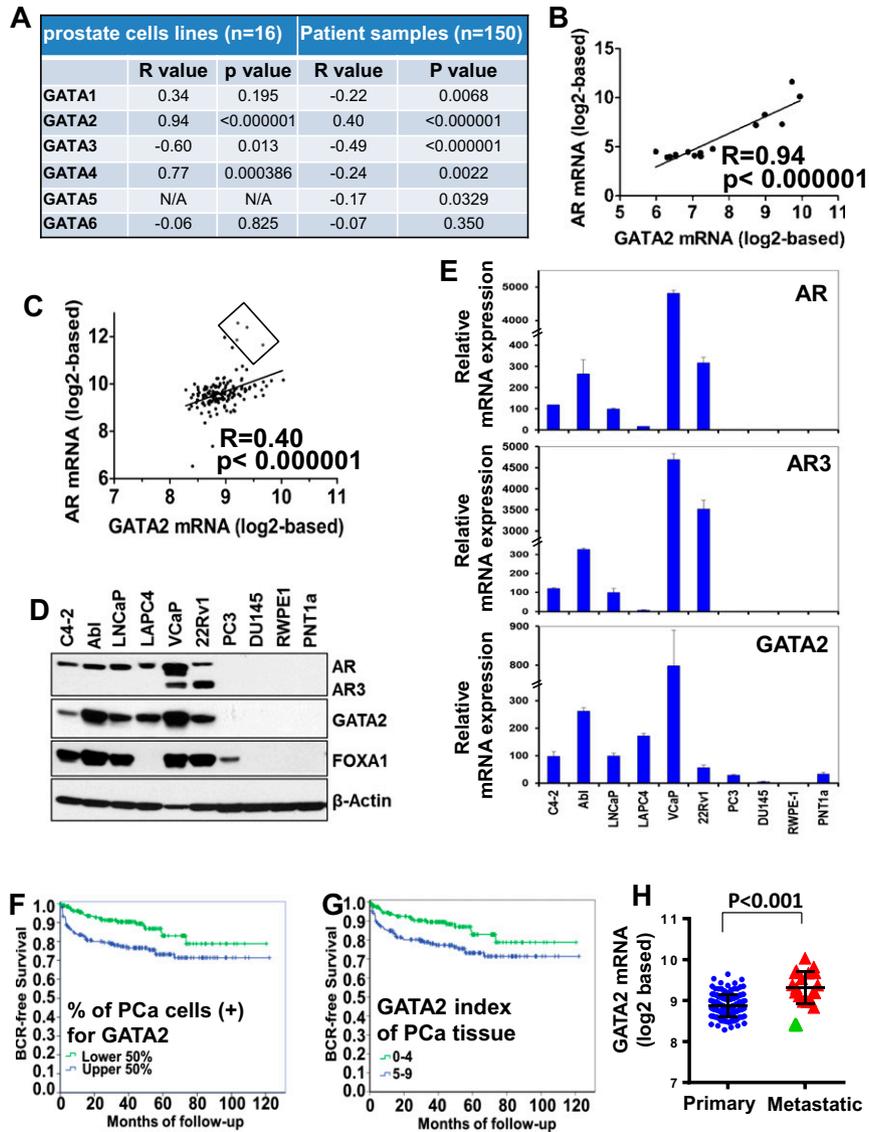
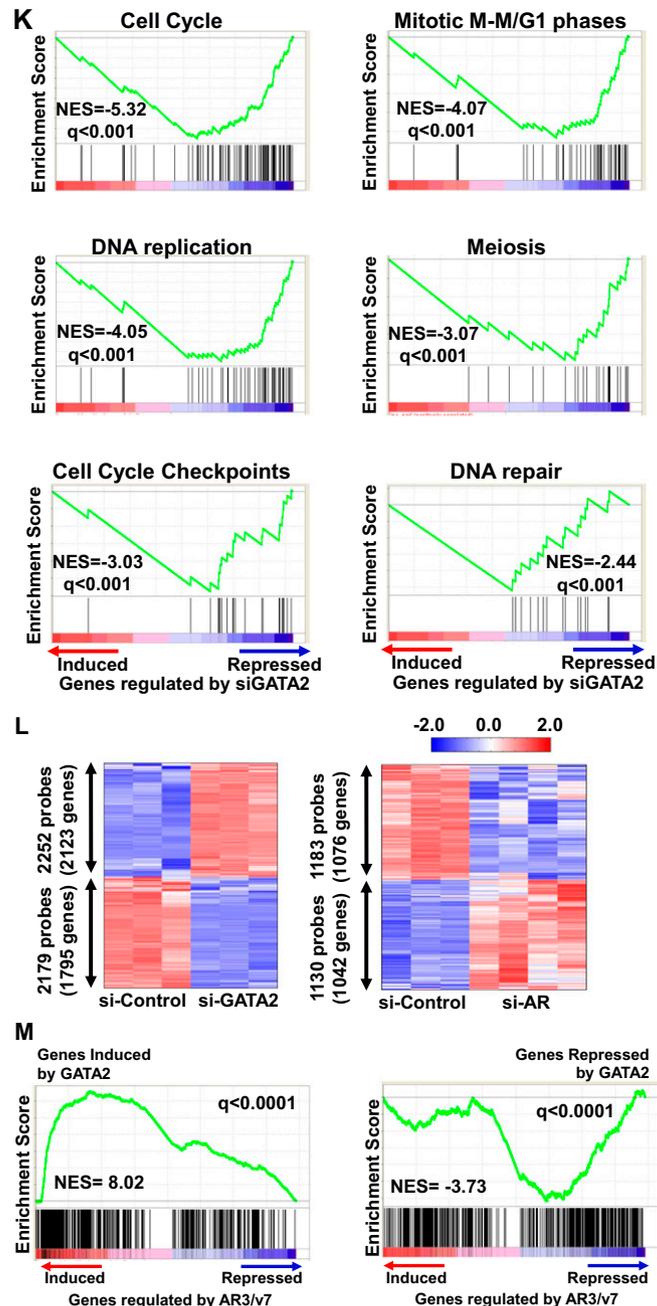


Fig. S1. (Continued)





**Fig. S1.** GATA2 expression is associated with AR expression, cell proliferation, and aggressive clinical behavior in PC. (A–E) Strong positive correlation of GATA2 with AR expression. We analyzed publicly available gene-expression datasets from a panel of 16 PC cell lines (GEO accession no. GSE9633) (35) and from 150 PC specimens (9). (A) The linear correlation between mRNAs of AR and the GATA2 family members (GATA1–6) was evaluated by calculation of the Pearson correlation coefficient. Among the six GATA2 family members, only GATA2 exhibited a positive and strong correlation with AR mRNA ( $r = 0.94$ ,  $P < 10^{-6}$  for the cell lines and  $r = 0.4$ ,  $P < 10^{-6}$  for the patient samples). (B) The positive correlation between AR and GATA2 mRNAs in the 16 PC cell lines is visualized in the scatterplot. (C) The positive correlation between AR and GATA2 mRNAs is visualized in the scatterplot (four samples with AR gene amplification are highlighted in a rectangle because they were, expectedly, outliers). (D and E) For validation, we performed immunoblot (D) and RT qPCR (E) analysis in 10 PC cell lines and confirmed that AR expression is restricted to GATA2<sup>+</sup> cell lines. (F) PCs in which a greater proportion of PC cells exhibited positive GATA2 immunostaining (the upper half of the cohort) had shorter BCR-free survival than the lower half of the cohort. IHC analysis for GATA2 was performed on a previously described tissue micro-array containing 383 clinically localized PCs. (G) PCs with a higher index of GATA2 immunostaining (upper half of the cohort) had shorter BCR-free survival than the lower half of the cohort. Compared with cancers with no or low expression (index 0–4,  $n = 192$ ), cancers with high levels of GATA2 expression (index >4,  $n = 191$ ) had a significantly decreased time to recurrence ( $P = 0.0072$ ). (H) GATA2 mRNA expression was higher in 19 metastatic specimens than in 131 primary PC specimens from a previously published cohort (9). The one metastatic specimen indicated by the green triangle represents a GATA2(-)/AR(-) basal prostate carcinoma that, expectedly, is an outlier. (I) The transcriptional response to GATA2 enriches for cell cycle pathways. Ingenuity pathway analysis revealed that the GATA2 gene signature (derived as described above by comparing samples treated with si-GATA2 with samples treated with si-NT) is highly enriched in transcripts involved in cellular growth and proliferation ( $P < 10^{-108}$ ), cell death ( $P < 10^{-93}$ ), and cell cycle ( $P < 10^{-88}$ ). (J) GSEA analysis of the GATA2 transcriptomic footprint vs. the REACTOME Pathways collection of the MSigDB shows strong enrichment of GATA2-regulated transcripts for cell cycle-related pathways (18 of the top 20 pathways,  $q < 0.001$  for all of them). (K) GSEA plots for select cell cycle pathways: cell cycle

Legend continued on following page

(NES = -5.32,  $q < 0.001$ ), mitotic M-M/G1 phases (NES = -4.07,  $q < 0.001$ ), DNA replication (NES = -4.05,  $q < 0.001$ ), meiosis (NES = -3.07,  $q < 0.001$ ), cell cycle checkpoints (NES = -3.03,  $q < 0.001$ ), and DNA repair (NES = -2.44,  $q < 0.001$ ) ( $q < 0.001$ ). (L) Hierarchical clustering of gene-expression profiles of LNCaP PC cells transfected with si-control or si-GATA2 (Left) or si-AR (Right) (genes differentially expressed,  $t$  test  $P < 0.05$ , fold change exceeding 4/3x). (M) GSEA analysis of a previously reported transcriptomic signature of AR3/v7 in LNCaP cells (12) vs. our GATA2 signature demonstrates that AR3/v7-regulated transcripts enrich positively for transcripts induced by GATA2 (i.e., down-regulated by si-GATA2, NES = 8.02,  $q < 0.0001$ ) (Left) and negatively for genes repressed by GATA2 (i.e., up-regulated by si-GATA2, NES = -3.73,  $q < 0.0001$ ) (Right).

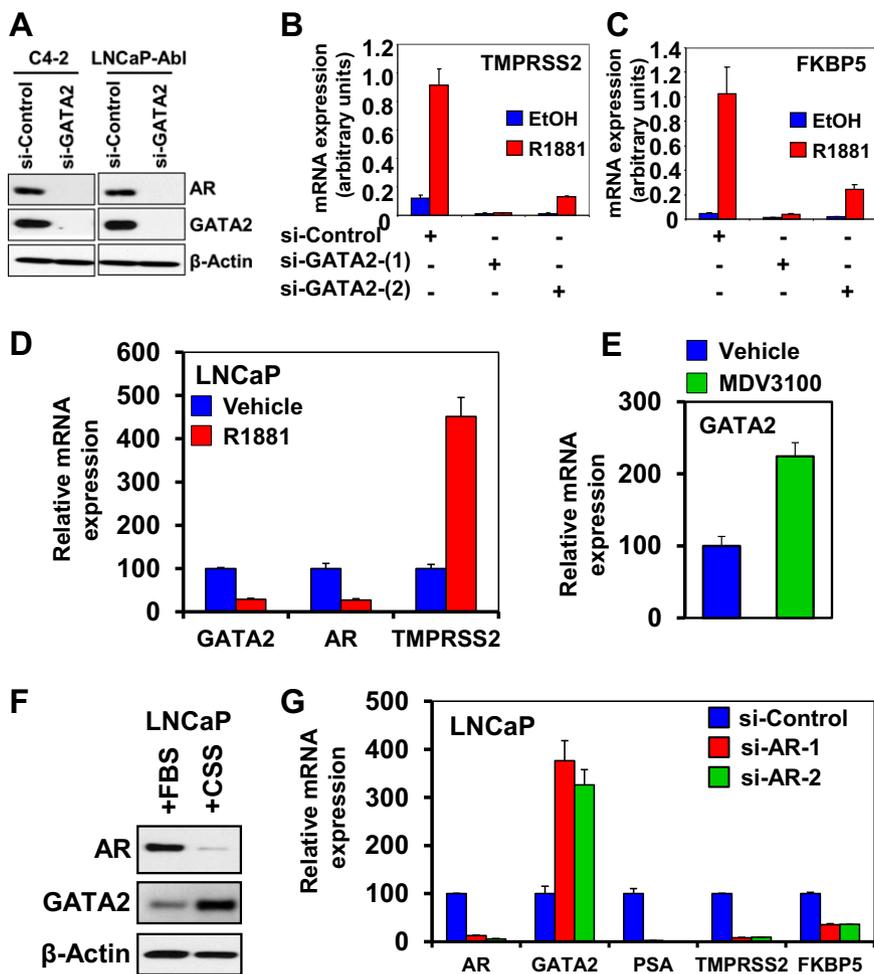


Fig. S2. (Continued)











**Table S1. Correlation of GATA2 expression (tumor index) with expression of other cancer-related proteins analyzed by quantitative immunohistochemistry in a tissue microarray of clinically localized prostate cancers**

Protein	Spearman's rho	P value
Androgen receptor	0.314	<0.0001
ERG	0.160	0.0493
SRC-1	0.227	0.0003
TIF2 (SRC-2)	0.144	0.0228
SRC-3	0.201	0.0309
Ki67	0.130	0.0394

Note: See references in text for prior publications of expression of each protein in this specimen cohort. The listed *P* values have not been adjusted for multiplicity.

**Table S2. Oligonucleotide sequences used in these studies**

Primer name	Forward primer 5'–3'	Reverse primer 5'–3'
AR promoter –5,472 bp	5'–GAATCTGCAGGCCAGTGTTT–3'	5'–AATCATCATGCAGGCCAAA–3'
AR peak +98.6 kb	5'–TTCCAATGGCATTTCATCCA–3'	5'–GCTTCGTGAGCCTCAAATCT–3'
AR peak +102.9 kb	5'–CCTTGTGATCATTGCTGGTA–3'	5'–CCCTCTTTGAAACATAATGAGGTA–3'
AR peak +133.3 kb	5'–AAAAAGAATGCACCGTACCC–3'	5'–AAAGATATTCAGCAGGTACGAAATG–3'
<i>KLK3</i> mRNA	5'–GTGCTTGTGGCCTCTCG–3'	5'–AGCAAGATCAGCCTTTTGTTC–3'
<i>TMPRSS2</i> mRNA	5'–CGCTGGCCTACTCTGGAA–3'	5'–CTGAGGAGTCGCACTCTATCC–3'
<i>FKBP5</i> mRNA	5'–GGATATACGCCAACATGTTCAA–3'	5'–CCATTGCTTTATTGGCCTCT–3'
GATA2 mRNA	5'–AAGGCTCGTTCCTGTTTCTCAGA–3'	5'–GGCATTGCACAGGTAGTGG–3'
AR mRNA	5'–AAGGCTATGAATGTCAGCCCA–3'	5'–CATTGAGGCTAGAGAGCAAGGC–3'
AR3/V7 mRNA	5'–CTTGTCTGCTTCGGAAATGTT–3'	5'–TCAGGGTCTGGTCATTTTGA–3'