



# Proteomic-Coupled-Network Analysis of T877A-Androgen Receptor Interactomes Can Predict Clinical Prostate Cancer Outcomes between White (Non-Hispanic) and African-American Groups

Naif Zaman<sup>1</sup>, Paresa N. Giannopoulos<sup>2</sup>, Shafinaz Chowdhury<sup>1,2</sup>, Eric Bonneil<sup>3</sup>, Pierre Thibault<sup>3</sup>, Edwin Wang<sup>1</sup>, Mark Trifiro<sup>2,4</sup>, Miltiadis Paliouras<sup>2,4\*</sup>

**1** Biotechnology Research Institute - National Research Council, Montréal, QC, Canada, **2** Lady Davis Institute for Medical Research, Segal Cancer Centre - Jewish General Hospital, Montréal, QC, Canada, **3** Institut de recherche en immunologie et en cancérologie, Université de Montréal, Montréal, QC, Canada, **4** Department of Medicine, Department of Oncology and Division of Experimental Medicine, McGill University, Montréal, QC, Canada

## Abstract

The androgen receptor (AR) remains an important contributor to the neoplastic evolution of prostate cancer (CaP). CaP progression is linked to several somatic AR mutational changes that endow upon the AR dramatic gain-of-function properties. One of the most common somatic mutations identified is Thr877-to-Ala (T877A), located in the ligand-binding domain, that results in a receptor capable of promiscuous binding and activation by a variety of steroid hormones and ligands including estrogens, progestins, glucocorticoids, and several anti-androgens. In an attempt to further define somatic mutated AR gain-of-function properties, as a consequence of its promiscuous ligand binding, we undertook a proteomic/network analysis approach to characterize the protein interactome of the mutant T877A-AR in LNCaP cells under eight different ligand-specific treatments (dihydrotestosterone, mibolerone, R1881, testosterone, estradiol, progesterone, dexamethasone, and cyproterone acetate). In extending the analysis of our multi-ligand complexes of the mutant T877A-AR we observed significant enrichment of specific complexes between normal and primary prostatic tumors, which were furthermore correlated with known clinical outcomes. Further analysis of certain mutant T877A-AR complexes showed specific population preferences distinguishing primary prostatic disease between white (non-Hispanic) vs. African-American males. Moreover, these cancer-related AR-protein complexes demonstrated predictive survival outcomes specific to CaP, and not for breast, lung, lymphoma or medulloblastoma cancers. Our study, by coupling data generated by our proteomics to network analysis of clinical samples, has helped to define real and novel biological pathways in complicated gain-of-function AR complex systems.

**Citation:** Zaman N, Giannopoulos PN, Chowdhury S, Bonneil E, Thibault P, et al. (2014) Proteomic-Coupled-Network Analysis of T877A-Androgen Receptor Interactomes Can Predict Clinical Prostate Cancer Outcomes between White (Non-Hispanic) and African-American Groups. PLoS ONE 9(11): e113190. doi:10.1371/journal.pone.0113190

**Editor:** Mohammad Saleem, Hormel Institute, University of Minnesota, United States of America

**Received:** May 10, 2014; **Accepted:** September 4, 2014; **Published:** November 19, 2014

**Copyright:** © 2014 Zaman et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The authors confirm that all data underlying the findings are fully available without restriction. All proteomic data presented in the manuscript, will be uploaded on our AR receptor database (<http://androgendb.mcgill.ca/>) as an excel file, together with the supplemental Figures/Tables associated with this manuscript.

**Funding:** This study was supported by the Canadian Institutes for Health Research (<http://www.cihr-irsc.gc.ca>) operating grant MOP-106477 (M.T., M.P., E.W.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* Email: [miltiadis.paliouras@mcgill.ca](mailto:miltiadis.paliouras@mcgill.ca)

## Introduction

Significant advances in genomic sequencing methodology have allowed a better assessment of the extent of somatic mutations accrued in common neoplasms [1,2]. More important is the realization that tumors significantly vary genetically from one patient to another and within a singular patient there exists extensive inter-tumoral heterogeneity and intra-tumoral heterogeneity [3,4,5]. A significant number of these genetic alterations are missense mutations that provoke new gain-of-function properties that render a particular gene proactive to tumoral evolution and are referred to as driver mutations. A better understanding of these new properties would lead to a better interpretation of oncogenesis, but this is difficult due to a large number of different

mutations, the unpredictable nature of gain-of-function properties associated with somatic mutations, the possible extensive interplay of different somatic mutants and the ensuing selection processes initiated by the microenvironment or by therapy itself. Such complex “systems” require a more global “omics” approach and more network analysis, rather than the classical single gene approach, to garner more critical information related to neoplastic evolution.

In keeping with the newly defined mutational landscape of tumors, prostate cancer (CaP) also has extensive genetic alterations that range from single missense mutations, copy number variation, splicing variants, genetic rearrangements and short DNA alterations in a large number of genes [1,2,6,7], including the androgen receptor (*AR*) gene. It is not unexpected that AR mutations can

add to the protein's repertoire of powerful new functions [8,9] and these gain-of-function attributes may allow the AR to function in an aberrant manner. A number of somatic CaP AR mutants, especially the most commonly occurring CaP AR mutation, Thr877Ala (T877A), have unique gain-of-function properties: they can bind several classes of steroids promiscuously (e.g. estrogens, progestins, glucocorticoids) with subsequent transactivation, or be hyperactivated by normal ligands [10]. Classic anti-androgen treatments [e.g. flutamide, cyproterone acetate (CPA) or bicalutamide] have generated, through selection pressure, specific somatic AR mutations, e.g. Trp741Cys (W741C) and His874Tyr, resulting in subversive ARs that are fully active with these drugs [11]. Even the next generation of anti-androgen drugs exemplified by enzalutamide (MDV-3100) has provoked specific AR mutations [12,13]. This observation also correlates with a dramatic fall in PSA levels subsequent to anti-androgen withdrawal [11]. The T877A-AR mutations, which is also present in prostate cancer cell line LNCaP, has been reported by various individuals to occur in 25 to 33% of androgen-independent or castrate-resistant tumors [11,14,15,16].

Recently, our own work strongly suggests that the AR function extends beyond its classical role as a transcription factor and includes the novel properties of RNA splicing, DNA methylation, proteasomal interaction and protein translation at the polyribosomes themselves [17]. Furthermore, the great functional diversity of the components of AR complexes exemplifies the intricate nature of protein-protein interactions associated with generating the appropriate AR biological output. These novel AR functions may mediate cellular processes and offer new areas in which somatic AR CaP mutants might “indulge” and promote CaP oncogenesis.

In an attempt to describe novel gain-of-function properties associated with mutant CaP ARs, a proteomic-coupled network analysis was performed. Multiple proteomics-mass spectroscopy investigations were carried out in order to fully characterize the protein composition of T877A-AR “complexes” (interactome) under different classes of hormone/ligand conditions reflecting the promiscuity of ligand binding associated with T877A. Critically, mutant CaP ARs may have their own unique ability to undergo and define new interactions. The coupling of the data generated by our proteomics screen to system biology analysis has been helpful in defining real and novel biological endpoints in AR complex systems, within a clinical disease perspective.

## Materials and Methods

### Cell lines

The LNCaP prostate cancer cell lines was obtained from the American Type Culture Collection (ATCC), Rockville MD.

### Cell culture, steroid hormones, ligands and stimulation experiments

LNCaP cell line was cultured in RPMI 1640 media supplemented with FBS (10%), at 37°C, 5% CO<sub>2</sub> in T-75 plastic culture flasks. Once confluent, the medium was changed to RMPI supplemented with 10% charcoal-dextran stripped FBS and incubated for an additional 24 h. The following day, the medium was changed to fresh RMPI/charcoal-dextran stripped FBS for overnight hormone/ligand stimulation studies for an 18 hour period. Steroids hormones were used at the following final concentrations, 10 nM dihydrotestosterone (DHT), 10 nM mibolerone (MB), 10 nM R1881, 10 nM testosterone+10 μM finasteride, 10 nM 17β-estradiol, 10 nM progesterone, 10 nM dexamethasone, and 100 nM cyproterone acetate (CPA).

### Affinity purification and Western blotting

LNCaP whole cell lysates were prepared by freeze-thaw method with 1X PDG buffer containing the appropriate hormone/ligand [18]. Lysates were then carried over for α-AR co-immunoprecipitations [AR(N20), Santa Cruz Biotechnology, Santa Cruz, CA] overnight at 4°C. A 50% Protein A Sepharose slurry was added to each sample and incubated at room temperature for 90 min. Beads were washed three times with wash buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% Tween 20) and resuspended in 100 μL 1X SDS gel loading buffer. Samples were denatured by boiling, and resolved on a 10% SDS-polyacrylamide gel before silver staining according to manufacturer's guidelines (BioRad) or transfer to a nitrocellulose membrane for Western blot analysis using monoclonal antibody AR(441) (NeoMarkers, Fremont, CA).

### Mass spectrometry and peptide comparison

TCEP (tris(2-carboxyethyl)phosphine) was added to the protein samples to reach the concentration of 5 mM. Samples were incubated at 37°C for 30 min. One μg of trypsin was added and the samples digested overnight at 37°C, then dried down in a SpeedVac and resolubilized in 50 μL of ACN 5%/formic acid (FA) 0.2%.

All MS analyses were performed using an LTQ-Orbitrap hybrid mass spectrometer with a nanoelectrospray ion source (ThermoFisher, San Jose, CA) coupled with an Eksigent nano-LC 2D pump (Dublin, CA) equipped with a Finnigan AS autosampler (Thermo Fisher, San Jose, CA). Twenty μL of each sample was injected on a C18 precolumn (0.3 mm i.d.×5 mm) and samples separated on a C18 analytical column (150 μm i.d.×100 mm) using an Eksigent nanoLC-2D system. A 76-min gradient from (A/B) 10–60% (A: formic acid 0.2%, B: acetonitrile/0.2% formic acid) was used to elute peptides with a flow rate set at 600 nL/min. The conventional MS spectra (survey scan) were acquired in profile mode at a resolution of 60,000 at m/z 400. Each full MS spectrum was followed by three MS/MS spectra (four scan events), where the three most abundant multiply charged ions were selected for MS/MS sequencing. Tandem MS experiments were performed using collision-induced dissociation in the linear ion trap.

The comparisons of peptide abundance across the different experimental paradigms were achieved using label-free quantitative proteomics [19,20]. Briefly, raw data files from the Xcalibur software was converted into peptide map files representing all ions according to their corresponding m/z values, retention time, intensity and charge state. Peptide abundance was then assessed using the “peak top” intensity values. Intensities of peptides eluting across several fractions were summed together, and only a coefficient of variance (CV) allowing the maximal ion transmission was considered to calculate peptide intensity. Clustering of peptide maps across different sample sets was performed on the peptide-associated Mascot entry using hierarchical clustering with specific tolerances (+/–15 ppm of peptide mass and +/–1 min of peptide retention time). Normalization of retention time was performed on the initial peptide cluster using a dynamic and nonlinear correction that confines the retention time distribution to less than 0.1 min on average. Reproducibility changes in abundance across conditions was determined using a two-tail homoscedastic *t*-test on sample replicates to identify peptide clusters with *p*-values <0.1 with fold changes greater than 7 standard deviations. Peptide clusters fulfilling these selection criteria was inspected manually to validate identification and changes in abundance. Expression analyses were performed on proteins identified by at least two different peptide sequences. Expression values and relative standard deviation were gained by averaging the intensity

differences and standard deviations of the four most intense peptide triplets after removing outlying peptide clusters. Normalized proteomic data can be found in **Table S1**.

### Datasets for network construction and gene expression analysis

All AR interactors were given NCBI gene IDs. Human protein interaction information was compiled from diverse data resources and annotation databases such as Biomolecular Interaction Network Database (BIND), the Database of Interacting Proteins (DIP), Human Protein Reference Database (HPRD), IntAct, and Molecular Interaction database (MINT), most of which contain curated interaction data and high-throughput data. We generated a metadata of protein interactions by merging these data with our own manually curated human signaling network containing 4,000 proteins and 22,000 signaling relations [21,22,23].

A protein interaction network was constructed for each hormone using our manually curated human signaling network and protein-to-protein interaction network. We only considered proteins that were  $\geq 2.5$  times the hormone condition abundance (signal) vs. the vehicle control abundance from the MS dataset, to be considered significantly present. In the network, a node and link represent the protein and interaction, respectively. To find the highly interconnected regions of the network, for each hormone, we scored each pair of interaction based on the number of neighboring nodes they have in common. This gave us a matrix with all the scores between all the interaction pairs. Hierarchical clustering was applied to the matrix and a threshold score calculated based on the partition density allowing us to identify the highly interconnected regions (clusters) for each of the hormone networks.

Next, we used these protein clusters and identified GO-terms (<http://www.geneontology.org/>) that are significantly associated with each of the protein clusters ( $p$ -value $<0.05$ , hyper-geometric). We also used Gene Set Enrichment Analysis (GSEA) defined pathways and performed GSEA if 25% of the protein cluster's genes were in the pathway. For the pathways and GO-terms that were significant from the GSEA results ( $p$ -value $<0.05$ ), we also did survival analysis. We used genes associated with these GO-terms and performed GSEA using GSE21034 [24].

### RNA extractions and microarray analysis

LNCAp cells were stimulated with panel of ligands as described above, and total RNA was extracted using TRIZOL (Invitrogen, Carlsbad, CA). RNA samples were then processed by Quebec Genome Innovation Centre (McGill University, Montreal, Canada), for microarray analysis with Illumina Human HT-12 Expression Beadchip v4 (Illumina, San Diego, CA). Raw data was processed using R [22,25].

Two global heatmaps were made. The first heatmap includes the most differentially expressed gene for each hormone using  $t$ -test ( $p$ -value $<0.05$ ), where each hormone is compared against all others. The second heatmap includes the top 10, 20, 50 and 100 genes with the highest variance. The  $t$ -test finds the most differentially expressed genes that are specific to each hormone, whereas the variance helps us observe genes that are differentially expressed across multiple hormones. GSEA analysis was performed using GSE21034 [24], of the 10, 20, 50 and 100 genes that showed the highest variance.

### Progression and Survival Analysis

Gene expression profiles, patient survival data, and demographic information for the 267 clinical prostate samples (29

normal, 181 primary and 37 metastatic tumors) were obtained from GSE21034 [24]. Breast, lung, lymphoma and medulloblastoma datasets were obtained from the Broad Institute (<http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi>) [26]. We examined the post-radical prostatectomy prognostic values of a subnetwork based on gene expression profiles of primary tumors, and performed Kaplan-Meier analysis by implementing the Cox-Mantel log-rank test using R as described previously [22,25]. If the  $p$ -value is less than 0.05, the subnetwork was treated as statistically significant to classify the tumors into non-metastatic and metastatic tumors. We stratified recurrent vs. non-recurrent CaP cancer based on the following criteria: PSA ( $\geq 4$  ng/mL), Gleason ( $\geq 7$ ), Tumor Stage ( $\geq 3$ ) and combined (PSA+Gleason+Tumor).

### Structure Preparation for MD simulation

The Crystal structure of DHT (1I38) and CPA (2OZ7) bound to T877A mutant AR-LBD and testosterone (2AM9) and R1881 (1E3K) bound to wild type AR-LBD are available in the Protein Data Bank (PDB). The complexes of different ligands bound to the T877A mutant AR-LBDs were further prepared for MD simulations using Xleap [27] in AMBER10 [28]. The generalized AMBER force field (GAFF) and ff99SB [29] parameters were used for eight different ligands. The complex was solvated in a truncated octahedron TIP3P [30] water box. The distance between the wall of the box and the closest atom of the solute was 12.0 Å, and the closest distance between the solute and solvent atoms was 0.8 Å. Counterions ( $\text{Cl}^-$ ) were added to maintain electroneutrality of the system. Each system was minimized, first by applying harmonic restraints with force constants of 10 kcal/mol/Å<sup>2</sup> to all solute atoms; second, by heating from 100 to 300 K over 25 ps in the canonical ensemble (NVT); and lastly by equilibrating to adjust the solvent density under 1 atm pressure over 25 ps in the isothermal-isobaric ensemble (NPT) simulation. The harmonic restraints were then gradually reduced to zero with four rounds of 25-ps NPT simulations. After additional 25-ps simulation, a 15-ns production run was obtained with snapshots collected every 1 ps. For all simulations, 2 fs time step and 9 Å non-bonded cut-off were used. The particle mesh Ewald method [31] was used to treat long-range electrostatics and bond lengths involving bonds to hydrogen atoms were constrained by SHAKE [32].

## Results

### Comparative network characterization of T877A-AR complexes: interactome and gene expression studies

Our ability to capture both ligand-bound and unliganded full-length wild-type AR by affinity chromatography under physiological conditions has previously allowed us to pursue a proteomics approach in order to characterize the components of wild-type AR complexes. This was done by subjecting such complexes to tryptic digestion followed by mass spectrometry (MS) to assign protein identification, in effect creating AR interactomes initiated by ligand binding [33]. To our MS data, a label-free quantitative method has now been applied across the different experimental paradigms (see Materials and Methods) [19,20], which allowed us to obtain data related to protein identification, along with abundance, thus allowing for direct comparisons between stimulation conditions.

The aim of differential hormone stimulation conditions will allow us to determine whether disease etiology of the T877A-AR mutation is dependent upon ligand and co-factor status. Therefore, we used LNCAp cells that endogenously express the T877A-AR mutation to characterize the ligand promiscuous protein

interactome complexes under different hormonal conditions, in order to highlight the possible distinct complexes that may be linked to disease progression. LNCaP cells were stimulated with the following hormones, four androgens: DHT, mibolerone (MB), R1881, or testosterone (in the presence of finasteride (to prevent the conversion of testosterone to DHT), or  $17\beta$ -Estradiol, progesterone, dexamethasone, or the anti-androgen cyproterone acetate (CPA), alone or with no-ligand/alcohol-vehicle control. Hormone stimulated T877A-AR complexes were immunopurified with an N-terminal specific AR antibody, which would not interfere with hormone binding to the AR ligand-binding domain. Eluates from all experimental conditions were analyzed by LC-MS/MS. In order to increase our sample frequency for peptide detection in our MS analysis, each experimental condition was performed four times. Our proteomics data is a compilation of only fully characterized proteins, with full gene ontology and function.

Quantitative MS data, for each of the eight hormone stimulation conditions, was used to create a protein interaction map (**Figure 1A**). The protein interaction network map, allows for a visual analysis of the relationship of the interaction of each protein with the mutant AR. It is most likely that not all proteins interact directly with the mutant AR, but can through intermediate proteins. Further ontological function classification (see Materials and Methods) is based on this interaction network map, by discerning significant clusters of interacting proteins based on the number of protein-protein interaction connections. Of the eight hormone stimulation T877A-AR protein lists, we then systemically applied hierarchical clustering analysis to the experimental conditions. Hierarchical clustering heat-maps representing the grouping of between whole T877A-AR agonist and antagonist experimental treatments were then generated (**Figure 1B**). Between the different experimental conditions (hormone treatments), a comparative network analysis was applied [21,22,23], and although four different androgens were used (DHT, testosterone, MB and R1881), the proteomic profiles of these androgen ligands do not segregate together, and we observed that progesterone and dexamethasone AR complexes have proteomic profiles that look like R1881 and MB, respectively. Moreover, the protein interaction complex for AR-estradiol-stimulated complexes was most similar to the AR-DHT response interactome.

We also proceeded to characterize the gene expression patterns of the multi-panel hormone stimulated LNCaP cells. Analysis of most variably expressed genes between the hormone conditions gave a hierarchical clustering pattern that was much different to the T877A-AR protein-interaction profile (**Figure 1C**). Quite clearly, we again observed that the different androgens used in these stimulation profiles do not segregate together, and that synthetic androgens, like R1881 and MB, do not have the same AR-stimulated transactivation profiles as the natural ligands like testosterone and DHT. Moreover, the functional ontological properties between protein-interaction vs. gene expression profiles of our differential ligand stimulated cells also appear to be very different. Discerning the impact on disease progression from these profiles was of particular interest.

To establish statistically significant biological functions, we implemented the incorporation of Gene Ontological (GO)/pathways terms using DAVID (Database for Annotation, Visualization and Integrated Discovery, <http://david.abcc.ncifcrf.gov/>). Using the protein interaction data from all the ligand stimulation conditions, the major ontological functions are: RNA pol II-dependent transcription, protein biosynthesis, with components of the translational machinery (translation initiation, elongation factors, ribosomal proteins and other regulatory proteins), RNA

metabolism (specifically RNA splicing), DNA repair (through an interaction with members of the DNA repair complex), and the proteasome/ubiquitination pathways (see **Table S2**). In contrast, the ligand-dependent gene expression pattern ontological classes included pathways involved in DNA replication, steroid/sterol biosynthesis, and apoptosis (see **Table S3**). Thus, although the AR is classically described as a transcription factor, its proteome profile would suggest that the AR is capable of functions beyond what has initially been described as a gene activator.

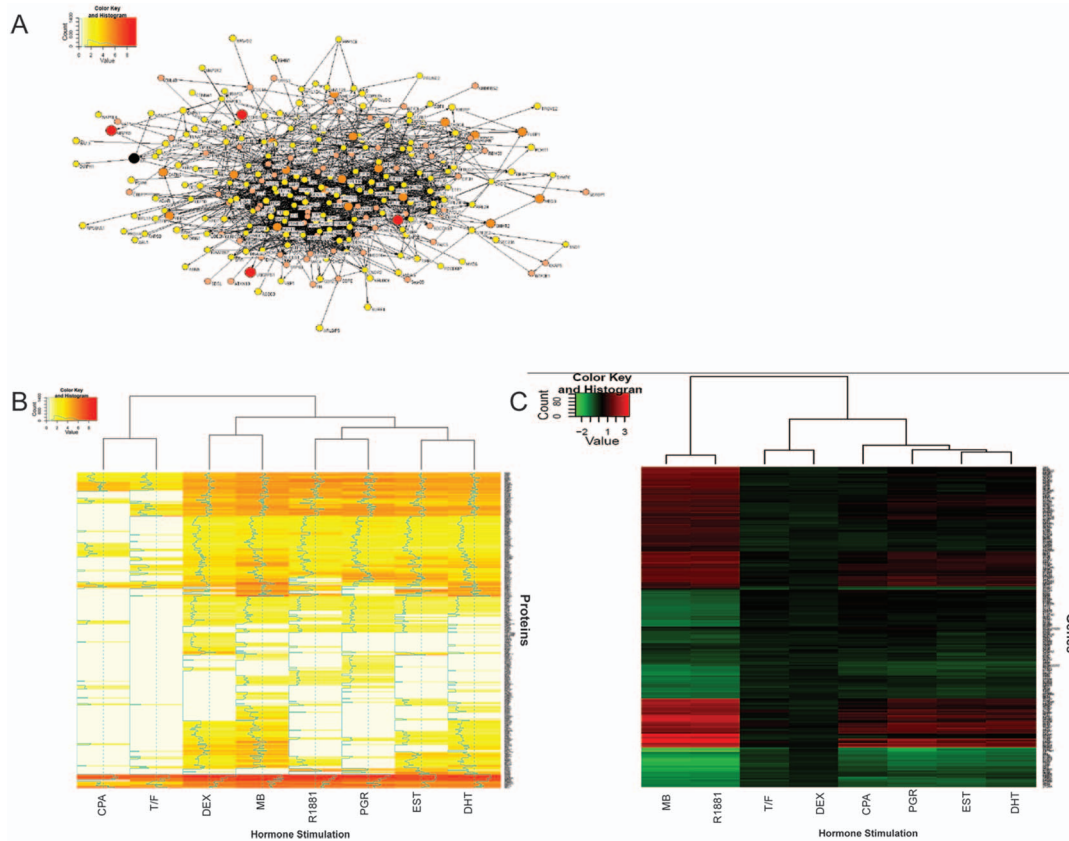
### Structural analysis of hormone binding to the T877A-AR

To explore the possible mechanisms by which ligands binding to the mutant AR create different sets of AR interacting proteins, we obtained the detailed conformation of the receptor, using 15 ns molecular dynamic (MD) simulation studies (see Materials and Methods) of the eight different ligands used. Using the docking program WILMA (**Figure 2**), we obtained structural data of the mutant AR bound with progesterone, estrogen, dexamethasone and MB. The structural data of the mutant AR with testosterone, DHT, R1881 and cyproterone acetate are already available and as such, these structures served as the starting points for the MD simulation studies.

Mutant AR MD simulations were performed over 15 ns production runs. To inspect the local flexibility of each protein/hormone complex, we calculated the root-mean-squared deviation (RMSD) fluctuations of backbone atoms of each amino acid residue for each mutant AR complex. In the study of globular protein conformations, one customarily measures the similarity in the three-dimensional structure by the RMSD of the central carbon atoms in amino acids, after optimal rigid body superposition. The most significant fluctuations correspond to loop regions between  $\alpha 3/\alpha 4$ ,  $\alpha 9$ ,  $\alpha 10/\alpha 11$  and  $\alpha 11/\alpha 12$  and the helices  $\alpha 11$  and  $\alpha 12$  themselves of the LBD of AR (**Figure 3**). It can be seen that the loop between  $\alpha 9$ ,  $\alpha 10$  and  $\alpha 11$  is the most flexible region in all the complexes (**Figure 3B**). Among the eight different AR ligand bound complexes, the testosterone- and estradiol-bound complexes demonstrate the highest flexibility, with some loop residues having RMSD fluctuations as high as 2.4 Å. Although these loops are distant from the hormone binding pocket, they are exposed to the surface and may serve a role as potential binding sites to other protein partners.

The other major differences observed between the mutant AR complexes are the positions of  $\alpha 11$  and 12, which are known to be critical for dictating hormone binding and co-activator interactions. Residues of  $\alpha 11$  and  $\alpha 12$  and the loop between them showed higher flexibility (**Figure 3C**). The T877A-AR mutation located in  $\alpha 11$  allows for a more spacious hormone-binding pocket and will accommodate steroids with different extensions within the D ring.

The examination of the nature and size of the solvent accessible surface area (SASA) of proteins is an important tool to measure potential interaction propensity with neighboring proteins. We calculated the average SASA of each AR complex from the 15 ns MD trajectory. No large differences were observed among the calculated SASAs of different AR mutant complexes, which ranged from 12,124 to 12,390 Å<sup>2</sup>. These results show that the differences are mostly associated with the  $\alpha 11$ - $\alpha 12$  regions of the AR-LBD, where the T877A mutation is located. Therefore, we decided to compare the dynamics among the eight different complexes, specifically in this region, by using RMSD matrices (**Table 1**). The matrices, which essentially capture extreme movements, reveal regions of high flexibility, especially for CPA and dexamethasone, compared to other AR-ligand complexes.



**Figure 1. Characterization of T877A-AR protein interaction and gene expression profiles.** **A.** Quantified protein interaction network of DHT stimulated LNCaP cells. The value of each protein, defined by our label-free quantitative MS, is distinguished by both color and size, and subsequent protein-protein interactions. The AR is designated by the black circle. To determine the relationship between protein interaction and gene expression patterns, hierarchical clustering of multi-panel hormone-stimulated LNCaP cells was carried out. **B.** Interacting proteins identified by mass spectrometry. **C.** Most variably expressed genes upon ligand stimulation. Although it has been suggested that all hormones used are able to activate androgen-dependent gene transcription with the T877A-AR mutant receptor, there are differences between AR protein complexes and AR gene expression patterns. This cluster analysis illustrates that even the synthetic androgens like Mibolerone (MB) and R1881, have similar gene expression profiles, their protein-interaction complexes are more similar to dexamethasone (DEX) and progesterone (PGR), respectively, than to either natural androgens testosterone and DHT. Moreover, even natural androgens (DHT and testosterone) can be segregated by their protein complexes. This would suggest that there are functions for the AR beyond gene transactivation. Protein and gene expression values are given as a ratio of quantified protein of each stimulation vs. vehicle control stimulation.  
doi:10.1371/journal.pone.0113190.g001

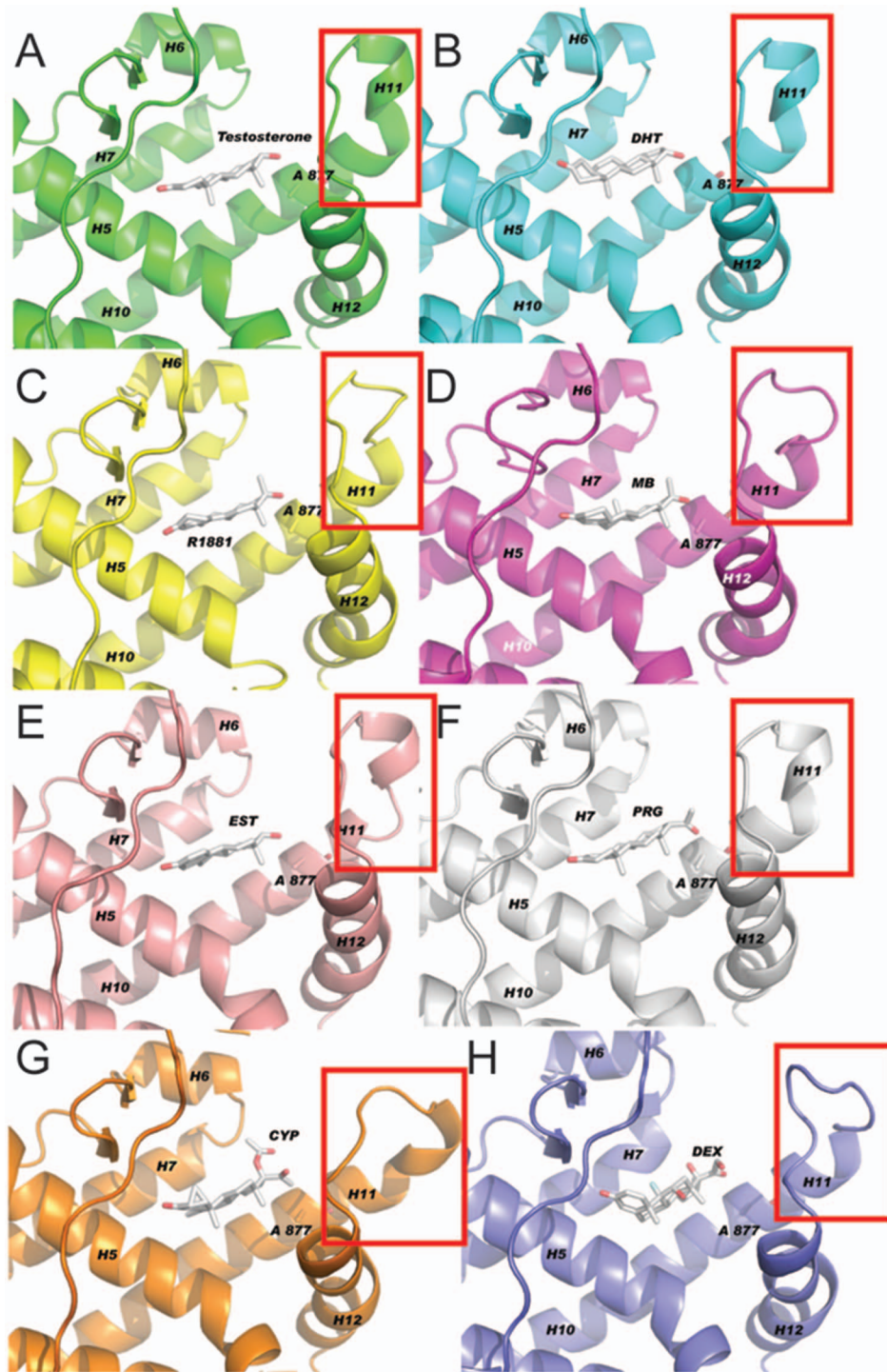
Computational modeling has also been carried out for a number of other AR somatic mutations in the ligand-binding domain (LBD), and show sensitivity to a broad range of hormone ligands, including, AR-W741L [34], -L701H [35,36,37,38], -H874Y [39,40,41] and -F876L [13]. Determination of the LBD structure between AR-WT and -H874Y (present in 22Rv1 cells), when bound to Testosterone in the presence of the N-terminal FXXLF motif peptide or the TIF2 coactivator peptide, found that all structures conformed to the canonical nuclear receptor LBD fold [41]. Moreover, the -H874Y DHT and R1881 structures conformed to -T877A and -W741L LBD bound to steroid and nosteroid ligands [34,42]. The double AR-mutant cell lines MDA-PCa-2a and MDA-PCa-2b, possess the L701H and T877A somatic mutations, these cells shows similar AR transactivation and LBD structural properties to the single AR-T877A mutant LNCaP cells to a broad spectrum of steroid ligands and anti-androgens, but also show an increased sensitivity cortisol steroids [35,36]. Most recently, the structure AR-F876L mutations has been investigated that allows the cell to use the new anti-androgen enzalutamide as an agonist. Similarly the Leu876 mutation allows for the antagonist-agonist switch, by accommodating of enzaluta-

mid to the ligand binding pocket [39]. Although only LBD structural data is available for the AR-F876L mutant, we don't believe that enzalutamide vs. androgen binding would significantly alter helices movement to drastically affect overall global AR structure, as has been observed with other AR LBD mutants.

#### AR protein interaction functional clusters, but not clinically derived gene expression profiles, correlate with CaP progression and survival outcome

From each of the ligand mutant AR protein interaction network, we identified specific sub-network modules. These sub-networks suggest hormone-specific activated pathways involved in either tumor initiation or progression. The characterization of ontological functions across the stimulation conditions is important as these differences or similarities in interacting proteins within the T877A-AR mutant complex may account for unique or shared cellular properties contributing to disease progression and outcomes. Therefore, we annotated significant GO-terms directly on sub-network modules extracted from the eight different hormone-protein interaction networks to highlight functions that





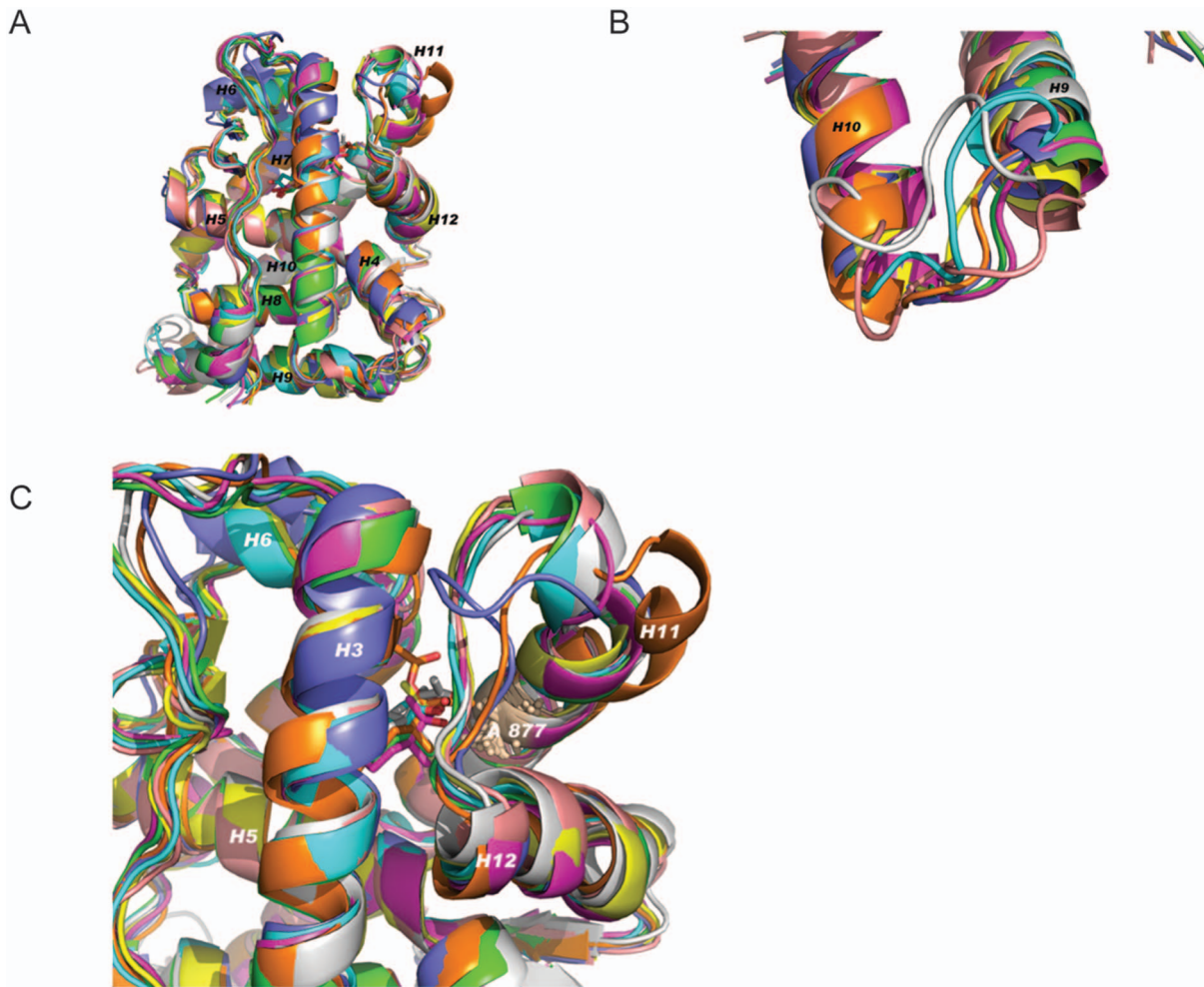
**Figure 2. MD average structure of T877A mutant of AR ligand-binding domain with eight different ligands.** Docking program WILMA was used to examine structural changes of the ligand-binding domain of the T877A mutation, upon binding to ligand binding. Illustrated is the average structure of T877A-AR mutation ligand binding domain with eight different ligands. **A.** testosterone, **B.** DHT, **C.** R1881, **D.** MB, **E.** estrogen (EST), **F.** progesterone (PRG), **G.** CPA, **H.** dexamethasone (DEX). A red box highlights the changes in the helix  $\alpha$  11 loop upon binding to each hormone ligand.

doi:10.1371/journal.pone.0113190.g002

may be unique to each of stimulation condition, and extracted the list of genes corresponding to those GO-terms.

Using the lists of genes from the annotated GO-terms (now to be referred to as “gene-sets”), we determined whether these gene-sets are enriched in the publicly available clinical prostatic tumor

microarray dataset, by applying Gene Set Enrichment Analysis (GSEA). We used the clinical data set GSE21034 [24], containing 247 clinical specimens (29 normal, 181 primary and 37 metastatic tumors). Initial analysis of all primary tumors from this data-set did not yield obvious enrichment of any gene-sets. However, after



**Figure 3. Loop structure fluctuations.** **A.** Superimposed average structures of all eight ligand bound receptor complexes. **B.** Regions of T877A AR-LBD loop between Helix $\alpha$ 9 and Helix $\alpha$ 10 showed maximum flexibility. **C.** Expanded view of Helix  $\alpha$ 11 and Helix  $\alpha$ 12 regions of T877A AR-LBD bound to eight different ligands studied in this study. Residues of  $\alpha$ 11 and  $\alpha$ 12 and loop between them showed higher flexibility compared to other regions of the receptor where T877A is located. Color corresponds to the following ligands: Green - testosterone, Cyan- DHT, Yellow-R1881, Pink- MB, Rose- EST, Grey- PGR, Orange-CPA, Purple-DEX. doi:10.1371/journal.pone.0113190.g003

further inspection of the data, unique features were noted in certain tumor samples. Thus, upon returning to the patient pathology information that accompanied the clinical data-set, two diverse patient populations could be immediately discerned, and thus we resegmented our data-sets between 142 White (non-Hispanic) and 25 African-American samples [other population groups (Hispanic and Asian) in the 181 available primary samples were too few to perform GSEA]. By segregating the dataset along available ethnic demographical information, we immediately were able to distinctly differentiate gene-sets between White (non-Hispanic) *vs.* African-American populations. From these results, we identified 138 T877A-AR -interacting protein sub-network modules (gene-sets) that show significant ( $p \leq 0.05$ ) enrichment of the T877A-AR -interacting partners that discerned CaP primary tumors *vs.* normal samples. Two gene-set examples are shown in **Figure 4**.

**Figure 4A** represents a gene-set (gene-set 1) that is significantly enriched among primary tumors of the African-American population, however subsequent GSEA of the same gene-set did not show significant enrichment in primary tumors of White (non-Hispanic) group. One of the unique features of the African-

American gene-set 1 is that is enriched in normal rather than tumor samples. This suggests that these genes may have anti-tumorigenic properties or subsequent loss of the expression may contribute detrimentally to prostate disease. Of particular note, the genes represented in gene-set 1 are part of the transcription-dependent DNA repair pathway. **Figure 4B** shows another gene-set (gene-set 2) that is significantly enriched in primary tumor samples of White (non-Hispanic) males and not in the corresponding African-American cohort. Furthermore, the GSEA results of these two T877A-AR interacting protein gene-sets are also distinct, and not shared, between primary tumor data of population groups of White (non-Hispanic) or African-American males. However, African-American gene-set 1 was significantly enriched in metastatic tumor data of White (non-Hispanic) males (data not shown). Due to limited data available on metastatic tumors from the African-American demographic (2 data sets), and reciprocal analysis could not be performed. This data highlights that we can identify gene-sets that show population-specific distinctions from primary tumors, but also suggests that the molecular characteristics of AR function underlying disease

**Table 1.** Two dimensional RMSD matrices of the residues of Helix  $\alpha$ 11 and Helix  $\alpha$ 12 and loop between them without fitting.

	DHT	TEST	MB	R1881	CPA	DEX	EST	PRG
DHT	0	1.72	1.83	1.79	3.11	3.23	1.79	2.03
TEST	1.72	0	1.69	2.05	3.17	3.19	1.73	2.2
MB	1.83	1.69	0	2.19	2.93	2.93	1.88	2.26
R1881	1.79	2.05	2.19	0	3.54	3.38	1.98	2.11
CPA	3.11	3.16	2.93	3.54	0	3.28	3.14	3.4
DEX	3.23	3.19	2.93	3.38	3.38	0	3.43	3.72
EST	1.79	1.73	1.88	1.98	3.12	3.46	0	1.82
PRG	2.01	2.18	2.26	2.11	3.41	3.72	1.83	0

doi:10.1371/journal.pone.0113190.t001

etiology in metastatic tumors may be common between African-American and White (non-Hispanic) men.

We next analyzed whether or not these distinct population T877A-AR gene-sets were predictive of 10-year survival outcomes, available from the GSE21034 data set [24], based on the following criteria; PSA ( $\geq 4$  ng/mL), Gleason ( $\geq 7$ ), Tumor stage ( $\geq T3$ ) or Combined (PSA + Gleason + Tumor stage). From our initial 138 characterized gene-sets, we identified 10 AR-interacting protein complexes that are indeed predictive for disease outcomes, 8 representing the White (non-Hispanic) population and the remaining the African-American group. From these 10 gene-sets (see **Table S4**), a single gene-set from the White (non-Hispanic) population (gene-set 2) was able to predict disease outcomes across all scoring criteria, (PSA, Gleason, Tumor Stage, and combined (**Figure 4C**)). This particular gene set did not exhibit any predictive value when analyzed against the African-American cohort of samples. The African-American gene-set (gene-set 1) was only able to predict disease outcome using the combined criteria (PSA+Gleason+Tumor Stage) (**Figure 4D**). All other gene-sets, from all population groups, were able to predict survival outcomes for only one scoring criteria (see **Figure S1**). Furthermore, these predictive clusters were shared between all the hormones, supporting the results from our dynamic modeling study, where the T877A mutation accommodates all steroid hormones to and exhibits very subtle structural differences, although the overall structures appear to elicit the same functional interaction platform.

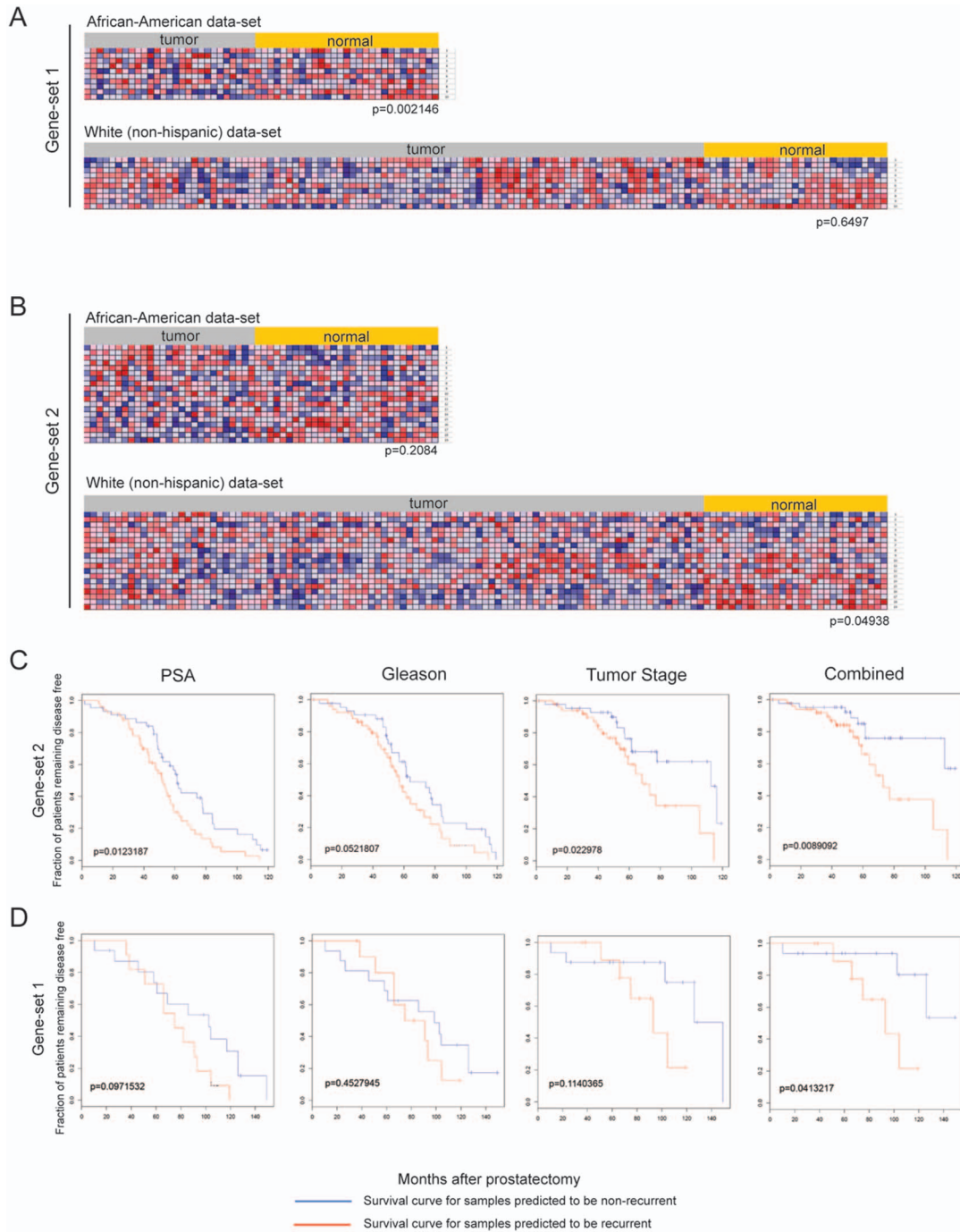
A similar analysis to that performed with our proteomic data was performed using our LNCaP multi-panel hormone microarray gene expression data across our hormone stimulation conditions. We selected 10, 20, 50 and 100 of the most variably expressed genes from our microarray data set to assess the ability to predict disease progression and outcome between White (non-Hispanic) or African-American men. We identified two gene-sets of 10 and 50 genes respectively that were able to distinguish between normal *vs.* tumor in White (non-Hispanic) men, but not African-American men. There appears to be no predictive value associated with each different class of hormone stimulation, irrespective of whether the hormones act through T877A-AR or through their cognate receptor. It is also apparently clear that the two different data-sets (protein interactome *vs.* gene expression), result in two different capabilities of predicting disease outcomes. This is a direct result of linking ontological function to a specific protein sub-network *vs.* arbitrarily selecting a defined number of genes linked solely to expression profiles.

Finally, although LNCaP cells are derived from a metastatic CaP lymph-node biopsy from a White (non-Hispanic) male, other cell lines also possessing the T877A-AR mutation, MDA-PCa-2a and MDA-PCa-2b, from bone metastatic CaP from African-American also exist. However, extensive genome-wide gene expression characterization between these cell lines and LNCaP, have found them to be most similar to one another *vs.* other androgen sensitive cell lines, LAPC4 (possessing a wild-type AR) or 22Rv1 (H874Y-AR) or *vs.* AR-null cell lines PC3 and DU145 [43]. Thus, if this differential hormone stimulation experiment were to be performed using MDA-PCa-2a or -2b cell lines, we would identify the same AR protein complexes *in vitro*, as LNCaP, and would also predict disease survival *in vivo* from clinical population data.

### Characterized T877A-AR protein gene-sets cannot predict survival outcomes in four other non-CaP cancers

We subsequently determined whether these predictive gene-sets were cancer specific and extracted gene expression datasets with available clinical outcome profiles for breast, lymphoma, lung and





**Figure 4. Gene expression profiles of AR protein sub-networks correlate to CaP progression and outcomes.** GSEA was applied to determine the correlation of transcript expression profiles of AR subnetwork protein clusters from raw microarray data of normal and primary tumor datasets of clinical prostatic samples (GSE21034). **A.** Illustrated is an enrichment profile of a single AR subnetwork which has significant differential expression between normal and tumor datasets, and was only significant for a population of White (non-Hispanic) but not African-American men. **B.** Enrichment profile for African-American men gene-set 1. Survival outcome analysis was followed-up on these gene-sets. Disease-free survival of CaP patients stratified by serum PSA, Gleason score, Tumor stage, and combined risk (PSA, Gleason and Tumor Stage) with respect to the AR subnetwork cluster. Kaplan-Meier analysis was used to plot the fraction of at-risk patients remaining free of disease (y-axis) at the indicated time after radical prostatectomy (x-axis). Patient stratification is based on serum PSA ( $\geq 4$  ng/mL), Gleason score ( $\geq 7$ ), Tumor Stage ( $\geq T3$ ) and combination of serum PSA, Gleason score and Tumor Stage score values. **C.** White (non-Hispanic) functional cluster gene-set 2. **D.** African-American functional clusters gene-set 1. doi:10.1371/journal.pone.0113190.g004

medulloblastoma clinical samples [26]. The gene-sets illustrated in **Figure 4** did not give significant outcome values for any of the four other cancers (**Figure 5**). Furthermore, the remaining 8

gene-sets, described above, also lacked significant predictive outcomes for the same 4 non-prostate cancers analyzed (see **Figure S2**). In hormone-dependent breast cancer, certain ethnic

population differences have been observed, with higher incidences of breast cancer occurring in African-American woman vs. other groups [44,45,46], but similar population demographic data used for the CaP cohort analyzed within our study was not available for the non-CaP cancers used in this analysis. Such genetic expression data would have been useful for further confirmatory follow-up studies. However, the expression of the AR has been described in a number of non-CaP cancers, especially breast cancer [47,48,49,50,51,52,53]. It has also been shown that several cell lines from these non-CaP cancers show androgen sensitivity and androgen-dependent gene expression profiles similar to CaP cell lines. However, we have identified AR interactome gene-sets that can differentiate between CaP and non-CaP disease survival which suggests that there are unique molecular characteristics of AR function in CaP and part of a CaP-specific pathway in neoplastic development, and also that these gene-sets can be used to predict CaP disease outcomes between genetically diverse groups.

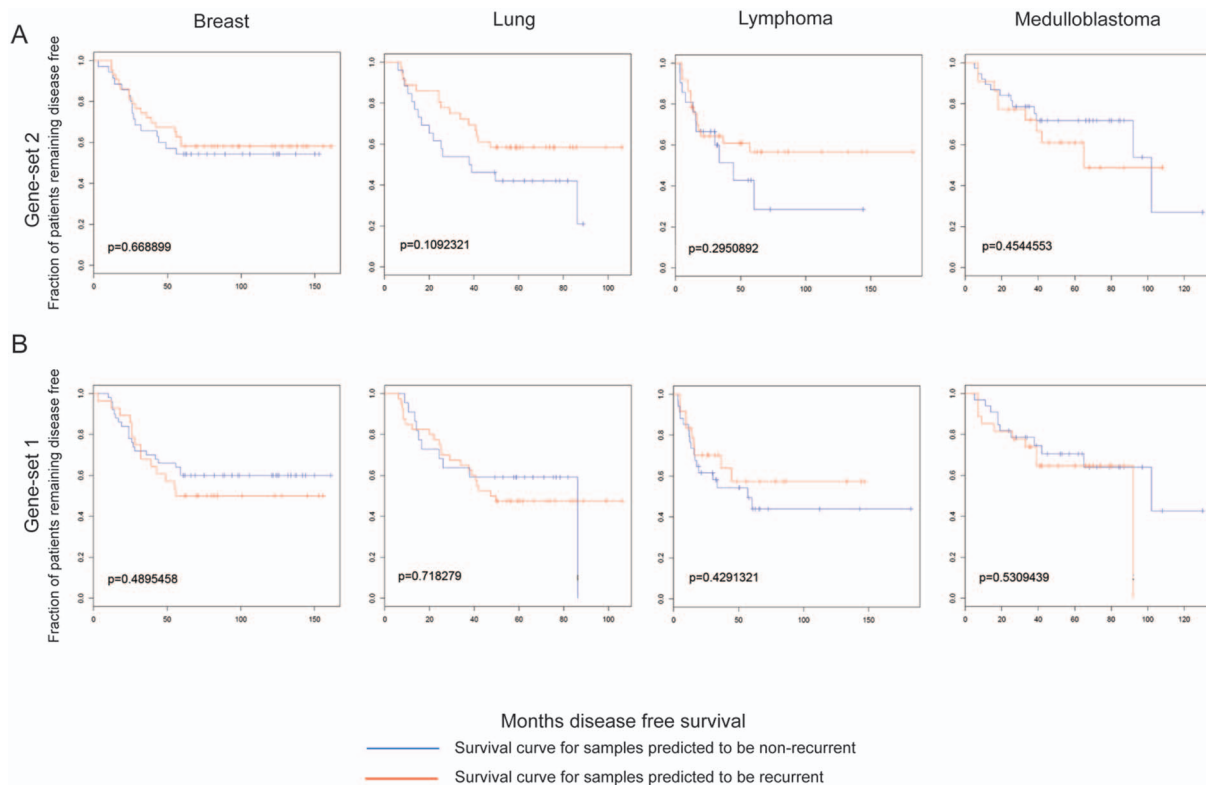
### White (non-Hispanic) gene-set can predict CaP disease outcome based on gene copy number

Finally, to deduce a molecular mechanism to account for differential gene expression patterns for each gene-set, we analyzed copy number variation of a CGH array dataset (GSE21035) [24], for the patients used for our GSEA and survival prediction outcome. For the White (non-Hispanic) gene-set 2, we were able to confirm that patient survival based on the criteria of PSA value, Gleason score and combined (PSA+Gleason), was dependent on copy number variation (**Figure 6**). This was the only gene-set that

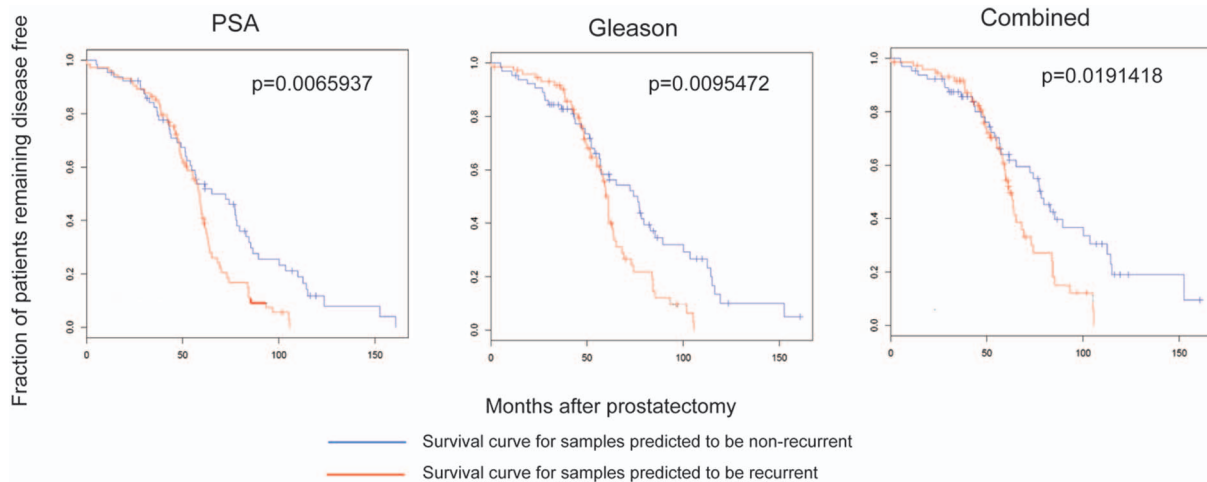
was predictive for survival outcomes based on copy number variation.

## Discussion

Knowledge of various molecular mechanisms of action contribute to our understanding of wild type AR function. Most mechanisms require the involvement of ligand binding and interacting partners. Examples of this include AR interactors involved in gene transactivation, including HSP70, HSP90, p300 and components of the RNA polII complex [54]. Gain-of-function somatic mutations, abundant in cancerous tissues, typically add new functions, adding to the complexity of physiological and disease outcomes. We investigated the T877A-AR mutation, as it represents the most common AR mutation in clinical CaP specimens, and is the AR mutation found in the most studied prostatic cancer cell line, LNCaP. Mutations like T877A-AR, and several others in the ligand-binding domain of the receptor, allow the AR to bind to other classes of steroid ligands such as estradiol, dexamethasone and progesterone, including anti-androgens such as CPA, resulting in subsequent AR dependent gene transactivation [55]. It is also now clear that as cancers evolve through many somatic mutations [56,57] and undergo selection processes induced by classical drug therapies themselves. Furthermore, a number of studies have used LNCaP cells as a model for studying the progression from androgen-dependent to -independent/castrate resistant prostate cancer (AIPC/CRPC) state and support the hypothesis that continuous AR activity and signaling continues to be one of the most important mechanisms in CRPC [58,59,60,61,62]. These studies have substantiated extensive



**Figure 5. Disease-free survival of 4 non-CaP cancers.** Datasets were retrieved from the Broad Institute (<http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi>) representing expression and disease outcomes for the following cancers: breast, lymphoma, lung and medulloblastoma, to determine survival outcomes for the gene-sets 1 and 2, as described in Figure 4. doi:10.1371/journal.pone.0113190.g005



**Figure 6. Survival outcomes reflected by Copy Number Variation.** Survival outcomes based on copy number variation (GSE21035) of gene-set 2 and same patients described in Figure 4, was applied. Patient stratification is based on serum PSA ( $\geq 4$  ng/mL) and Gleason score ( $\geq 7$ ) and combination of serum PSA and Gleason score values. doi:10.1371/journal.pone.0113190.g006

genetic alterations that range from single missense mutations, to copy number variations, splicing variants, genetic rearrangements and short DNA alterations in a large number of genes and AR co-factor interactions to reproduce androgen-independent scenario [1,2,6,7,36,63,64]. Modeling AIPC using LNCaP cell lines and actual tissue from AIPC patients, Wang et al., 2009 [64] found that the gene expression regulated by the AR in the absence of hormone is distinct from androgen-regulated program and can selectively and directly upregulate M-phase genes found in androgen-independent CaP and may explain why maximal androgen deprivation (AR antagonists and LHRH inhibitors), and cannot prolong androgen-independent survival. Most recently, it was found that overexpression of AR was a result to prolonged exposure of LNCaP-derived xenografts with the anti-androgen enzalutamide, and was similar to chronic androgen depletion, as a significant mechanism for drug resistance and CRPC development [65].

In this study we characterized whole mutant AR protein complexes with several classes of steroids and ligands known to bind T877A-AR. The specific interactomes were dependent on the ligand utilized; so too were the specific gene expression profiles associated with each ligand. Thus mutant AR gain-of-function properties are not singular but multiple, dictated by the class of steroid hormones used. Further exploration of other adrenal androgens such as DHEA or androstenedione or other anti-androgens such as flutamide and bicalutamide were not examined in this study, however, we did select a diverse class of ligands known to bind to the T877A-AR variant.

High-throughput gene expression microarray approaches described in CaP cells have identified hundreds of androgen-regulated genes and also characterized genome-wide AR recruitment sites [66,67,68,69]. The classical AR complex contains general transcription factors, coregulators and specific transcription factors that associate either directly or indirectly with the AR to enhance or repress its transcriptional activity function without themselves necessarily binding to DNA. As shown in our recent proteomic studies [17], including this one, AR complexes may also include a larger number of functionally diverse proteins involved in a multitude of “non-classical” AR cellular processes such as histone acetylation, DNA methylation, ubiquitination, RNA

splicing, apoptosis, and protein synthesis, with all pathways found to be dependent on hormone stimulation conditions.

From our data, there are several clusters of AR-interacting proteins that are worth exploring to understand their role in disease progression. The first cluster of AR-interacting proteins is unique to the African-American population group and consists of the following proteins: **ERCC1**, **ERCC2**, **ERCC3**, **ERCC5** and **FEN1**. The function of these proteins is required for mediating DNA damage excision repair. However, they are known components of the RNA polymerase II transcriptional complex [70] and ERCC2 and ERCC3 have been previously published as AR interactors [71]. Experimental evidence show that over-activity of this complex can lead to instability of CAG/CTG triplet repeats, resulting in a shortening of the repeat [72,73].

The second cluster of AR-interacting proteins is unique for the White (non-Hispanic) population and is involved in chromatin remodeling and histone deacetylation activity. From this cluster we identified the following proteins: **KDM1A**, **KDM4C**, **SIRT1**, **CTBP1**, **NR2C1** and **SMARCD3**. NR2C1 has been previously described as an AR interactor [74]. The selection of this cluster of proteins for further analysis would be interesting because of the increased attention histone deacetylase inhibitors that are garnering in cancer biology [75].

A final pathway for further investigation, and unique to the White (non-Hispanic) group of men, is the role of the AR in participating in the negative regulation of apoptosis via its interaction with **BCL2**, **RELA**, **FAS**, **EEF1A2** and **NR4A2**. The role of these proteins have been well described in apoptotic pathways, and RELA (NF $\kappa$ B p65) is a well characterized interactor of AR [76] and BCL2 [77]. A proposed mechanism by which these proteins may facilitate regulating apoptosis would be via a signal transduction cascade that would negatively regulate pro-apoptotic genes and proteins [77].

Using the T877A-AR hormone-specific interaction complexes as a basis for a novel systems biology network analysis exercise established gene-sets with clear predictive CaP clinical outcome value. In doing so, we confirmed the critical importance of the genetic backgrounds of the CaP individuals in the clinical dataset. Without segregating the microarray expression data of CaP patients along White (non-Hispanic) and African-American datasets, no defined gene-sets with predictive clinical values could

be identified. Once the data had been segregated, gene-sets with very powerful clinical outcome parameters were discovered. African-American men have long been considered to have clinically different CaP from White (non-Hispanic) men, based on their genetic background [78,79]. This is not entirely surprising, as African-American descent has long been associated with higher incidence and more aggressive disease, characterized by greater tumor volume for each clinical stage, have greater PSA levels and a more aggressive cancer for Gleason score of 8 or greater with compare to White (non-Hispanics) males [80,81]. Investigations, excluding socio-economic disparities that would limit individuals to health care [82], to explain racial differences between African-American vs. White (non-Hispanic) males, have excluded hormones levels, as serum testosterone levels (later in life) have similar levels at the time of prostate biopsy and in the their prostate biopsy tissue [83,84]. However, analysis of AR expression in malignant vs. benign prostate tissue African-American males found to be 27% more likely to stain positive for AR and the nuclear localization of the AR was 81% greater than White (non-Hispanics) [85]. It was also observed that there was significantly expression of CaP biomarkers that those of white men, one of which was the AR [86]. Interestingly, African-American males vs. other ethnic groups display shorter AR CAG repeat lengths which code for the polyglutamine tract of the AR [87]. AR with shorter polyglutamine tracts exhibits higher AR activity and represents a potential risk factor for CaP [88,89,90]. Therefore, dysregulation of DNA repair function, specifically loss of expression, and the link to contraction of CAG tract length in African-American males is a mechanism to investigate as a mechanism to explain racial differences in CaP. Furthermore, to gain more insight into CaP molecular/genetic etiology, our approach, encompassing proteomics, expression studies and network analysis, would benefit from investigating even more populations of diverse genetic origin backgrounds that have very low rates of CaP including Chinese and Middle Eastern men. By identifying such protective pathways of AR function, along with identifying powerful prognostic tools to predict disease, now we can assess pathways to also offer novel therapeutic targets.

## References

- Kumar A, White TA, MacKenzie AP, Clegg N, Lee C, et al. (2011) Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proc Natl Acad Sci U S A* 108: 17087–17092.
- Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, et al. (2011) The genomic complexity of primary human prostate cancer. *Nature* 470: 214–220.
- Barbieri CE, Tomlins SA (2014) The prostate cancer genome: Perspectives and potential. *Urol Oncol* 32: 15–22.
- Xu X, Zhu K, Liu F, Wang Y, Shen J, et al. (2013) Identification of somatic mutations in human prostate cancer by RNA-Seq. *Gene* 519: 343–347.
- Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, et al. (2013) Punctuated evolution of prostate cancer genomes. *Cell* 153: 666–677.
- Wu C, Wyatt AW, Lapuk AV, McPherson A, McConeghy BJ, et al. (2012) Integrated genome and transcriptome sequencing identifies a novel form of hybrid and aggressive prostate cancer. *J Pathol* 227: 53–61.
- Hieronymus H, Sawyers CL (2012) Traversing the genomic landscape of prostate cancer from diagnosis to death. *Nat Genet* 44: 613–614.
- Bielas JH, Loeb KR, Rubin BP, True LD, Loeb LA (2006) Human cancers express a mutator phenotype. *Proc Natl Acad Sci U S A* 103: 18238–18242.
- Venkatesan RN, Bielas JH, Loeb LA (2006) Generation of mutator mutants during carcinogenesis. *DNA Repair (Amst)* 5: 294–302.
- Vogelstein B, Kinzler KW (2002) The genetic basis of human cancer. New York: McGraw-Hill, Medical Pub. Division. xv, 821 p.
- Taplin ME, Bubley GJ, Ko YJ, Small EJ, Upton M, et al. (1999) Selection for androgen receptor mutations in prostate cancers treated with androgen antagonist. *Cancer Res* 59: 2511–2515.
- Joseph JD, Lu N, Qian J, Sensintaffar J, Shao G, et al. (2013) A clinically relevant androgen receptor mutation confers resistance to second-generation antiandrogens enzalutamide and ARN-509. *Cancer discovery* 3: 1020–1029.
- Korpai M, Korn JM, Gao X, Rakiec DP, Ruddy DA, et al. (2013) An F876L mutation in androgen receptor confers genetic and phenotypic resistance to MDV3100 (enzalutamide). *Cancer discovery* 3: 1030–1043.
- Gaddipati JP, McLeod DG, Heidenberg HB, Sesterhenn IA, Finger MJ, et al. (1994) Frequent detection of codon 877 mutation in the androgen receptor gene in advanced prostate cancers. *Cancer Res* 54: 2861–2864.
- Taplin ME, Bubley GJ, Shuster TD, Frantz ME, Spooner AE, et al. (1995) Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer. *N Engl J Med* 332: 1393–1398.
- Veldscholte J, Ris-Stalpers C, Kuiper GG, Jenster G, Berrevoets C, et al. (1990) A mutation in the ligand binding domain of the androgen receptor of human LNCaP cells affects steroid binding characteristics and response to antiandrogens. *Biochem Biophys Res Commun* 173: 534–540.
- Paliouras M, Zaman N, Lumbroso R, Kapogeorgakis L, Beitel LK, et al. (2011) Dynamic rewiring of the androgen receptor protein interaction network correlates with prostate cancer clinical outcomes. *Integrative biology : quantitative biosciences from nano to macro* 10: 1020–1032.
- Beitel LK, Sabbaghian N, Alarifi A, Alvarado C, Pinsky L, et al. (1995) Characterization of normal and point-mutated human androgen receptors expressed in the baculovirus system. *J Mol Endocrinol* 15: 117–128.
- Saba J, Bonnell E, Pomies C, Eng K, Thibault P (2009) Enhanced sensitivity in proteomics experiments using FAIMS coupled with a hybrid linear ion trap/Orbitrap mass spectrometer. *J Proteome Res* 8: 3355–3366.
- Kearney P, Thibault P (2003) Bioinformatics meets proteomics—bridging the gap between mass spectrometry data analysis and cell biology. *J Bioinform Comput Biol* 1: 183–200.
- Awan A, Bari H, Yan F, Moksong S, Yang S, et al. (2007) Regulatory network motifs and hotspots of cancer genes in a mammalian cellular signalling network. *IET Syst Biol* 1: 292–297.
- Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, et al. (2007) A map of human cancer signaling. *Mol Syst Biol* 3: 152.
- Cui Q, Yu Z, Purisima EO, Wang E (2006) Principles of microRNA regulation of a human cellular signaling network. *Mol Syst Biol* 2: 46.

In conclusion, our unique approach of using an important gain-of-function AR mutation, has generated gene-sets along functional organization lines showing that we can distinguish prostatic disease between White (non-Hispanic) and African-American men. Moreover, in our study for classification of AR function based on interaction profiles was a much more powerful tool predicting disease and survival outcomes than analyzing androgen-dependent gene expression patterns. The identification of these new functional properties of the AR and somatic mutations of the AR, further suggests that there is a role for the AR beyond that as a transcription factor and also implicates the ability of non-androgenic hormones to activate CaP disease-linked pathways.

## Supporting Information

### Figure S1 CaP 10-year survival outcomes for T877A-AR Gene-sets 3–10.

(PDF)

### Figure S2 Disease-free survival outcomes of 4 non-CaP cancers for Gene-sets 3–10.

(PDF)

### Table S1 Normalized Proteomic Data.

(XLS)

### Table S2 GO-Term definitions of Proteomic data.

(XLS)

### Table S3 GO-Term definitions of microarray data.

(XLS)

### Table S4 GO-Term definitions of T877A-AR Gene-sets 1–10.

(XLS)

## Author Contributions

Conceived and designed the experiments: MP MT EW. Performed the experiments: NZ PG SC EB. Analyzed the data: SC EB PT MP. Contributed reagents/materials/analysis tools: PT EW MT MP. Contributed to the writing of the manuscript: EB MT MP.



24. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, et al. (2010) Integrative genomic profiling of human prostate cancer. *Cancer Cell* 18: 11–22.
25. Li J, Lenferink AE, Deng Y, Collins C, Cui Q, et al. (2010) Identification of high-quality cancer prognostic markers and metastasis network modules. *Nature communications* 1: 34.
26. Ramaswamy S, Ross KN, Lander ES, Golub TR (2003) A molecular signature of metastasis in primary solid tumors. *Nat Genet* 33: 49–54.
27. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, et al. (2005) The AMBER biomolecular simulation programs. *J Comput Chem* 26: 1668–1688.
28. Gotz AW, Williamson MJ, Xu D, Poole D, Le Grand S, et al. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation* 8: 1542–1555.
29. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, et al. (2006) Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins* 65: 712–725.
30. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* 79: 926–935.
31. Darden T, York D, Pedersen L (1993) Particle mesh ewald - an N.Log(N) method for ewald sums in large systems. *Journal of Chemical Physics* 98: 10089–10092.
32. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of N-alkanes. *Journal of Computational Physics* 23: 327–341.
33. Parker CE, Warren MR, Loisele DR, Dicheva NN, Scarlett CO, et al. (2005) Identification of components of protein complexes. *Methods Mol Biol* 301: 117–151.
34. Bohl CE, Gao W, Miller DD, Bell CE, Dalton JT (2005) Structural basis for antagonism and resistance of bicalutamide in prostate cancer. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 102: 6201–6206.
35. Zhao XY, Boyle B, Krishnan AV, Navone NM, Peehl DM, et al. (1999) Two mutations identified in the androgen receptor of the new human prostate cancer cell line MDA PCa 2a. *J Urol* 162: 2192–2199.
36. Zhao XY, Malloy PJ, Krishnan AV, Swami S, Navone NM, et al. (2000) Glucocorticoids can promote androgen-independent growth of prostate cancer cells through a mutated androgen receptor. *Nat Med* 6: 703–706.
37. Matias PM, Carrondo MA, Coelho R, Thomaz M, Zhao XY, et al. (2002) Structural basis for the glucocorticoid response in a mutant human androgen receptor (AR(ccr)) derived from an androgen-independent prostate cancer. *J Med Chem* 45: 1439–1446.
38. van de Wijngaert DJ, Molier M, Lusher SJ, Hersmus R, Jenster G, et al. (2010) Systematic structure-function analysis of androgen receptor Leu701 mutants explains the properties of the prostate cancer mutant L701H. *J Biol Chem* 285: 5097–5105.
39. Mahmoud AM, Zhu T, Parray A, Siddique HR, Yang W, et al. (2013) Differential effects of genistein on prostate cancer cells depend on mutational status of the androgen receptor. *PLoS One* 8: e78479.
40. Zhou J, Geng G, Shi Q, Sauriol F, Wu JH (2009) Design and synthesis of androgen receptor antagonists with bulky side chains for overcoming antiandrogen resistance. *J Med Chem* 52: 5546–5550.
41. Askew EB, Gampe RT Jr, Stanley TB, Faggart JL, Wilson EM (2007) Modulation of androgen receptor activation function 2 by testosterone and dihydrotestosterone. *J Biol Chem* 282: 25801–25816.
42. Bohl CE, Wu Z, Miller DD, Bell CE, Dalton JT (2007) Crystal structure of the T877A human androgen receptor ligand-binding domain complexed to cyproterone acetate provides insight for ligand-induced conformational changes and structure-based drug design. *J Biol Chem* 282: 13648–13655.
43. Zhao H, Kim Y, Wang P, Lapointe J, Tibshirani R, et al. (2005) Genome-wide characterization of gene expression variations and DNA copy number changes in prostate cancer cell lines. *Prostate* 63: 187–197.
44. Wu AH, Gomez SL, Vigen C, Kwan ML, Keegan TH, et al. (2013) The California Breast Cancer Survivorship Consortium (CBCSC): prognostic factors associated with racial/ethnic differences in breast cancer survival. *Cancer causes & control: CCC* 24: 1821–1836.
45. Long J, Zhang B, Signorello LB, Cai Q, Deming-Halverson S, et al. (2013) Evaluating genome-wide association study-identified breast cancer risk variants in African-American women. *PLoS One* 8: e58350.
46. Lindner R, Sullivan C, Offor O, Lezon-Geyda K, Halligan K, et al. (2013) Molecular phenotypes in triple negative breast cancer from African American patients suggest targets for therapy. *PLoS One* 8: e71915.
47. Park HS, Bac JS, Noh SJ, Kim KM, Lee H, et al. (2013) Expression of DBC1 and androgen receptor predict poor prognosis in diffuse large B cell lymphoma. *Translational oncology* 6: 370–381.
48. Mikkonen L, Pihlajamaa P, Sahu B, Zhang FP, Janne OA (2010) Androgen receptor and androgen-dependent gene expression in lung. *Mol Cell Endocrinol* 317: 14–24.
49. Recchia AG, Musti AM, Lanzino M, Panno ML, Turano E, et al. (2009) A cross-talk between the androgen receptor and the epidermal growth factor receptor leads to p38MAPK-dependent activation of mTOR and cyclinD1 expression in prostate and lung cancer cells. *Int J Biochem Cell Biol* 41: 603–614.
50. Kaiser U, Hofmann J, Schilli M, Wegmann B, Klotz U, et al. (1996) Steroid-hormone receptors in cell lines and tumor biopsies of human lung cancer. *International journal of cancer Journal international du cancer* 67: 357–364.
51. McGhan IJ, McCullough AE, Protheroe CA, Dueck AC, Lee JJ, et al. (2014) Androgen receptor-positive triple negative breast cancer: a unique breast cancer subtype. *Annals of surgical oncology* 21: 361–367.
52. Vera-Badillo FE, Templeton AJ, de Gouveia P, Diaz-Padilla I, Bedard PL, et al. (2014) Androgen receptor expression and outcomes in early breast cancer: a systematic review and meta-analysis. *Journal of the National Cancer Institute* 106: djt319.
53. Ren Q, Zhang L, Ruoff R, Ha S, Wang J, et al. (2013) Expression of androgen receptor and its phosphorylated forms in breast cancer progression. *Cancer* 119: 2532–2540.
54. Heemers HV, Tindall DJ (2007) Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. *Endocr Rev* 28: 778–808.
55. Ngan S, Stronach EA, Photiou A, Waxman J, Ali S, et al. (2009) Microarray coupled to quantitative RT-PCR analysis of androgen-regulated genes in human LNCaP prostate cancer cells. *Oncogene* 28: 2051–2063.
56. Lee W, Jiang Z, Liu J, Haverty PM, Guan Y, et al. (2010) The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 465: 473–477.
57. Pleasance ED, Cheatham RK, Stephens PJ, McBride DJ, Humphray SJ, et al. (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463: 191–196.
58. Marques RB, Dits NF, Erkens-Schulze S, van Weerden WM, Jenster G (2010) Bypass mechanisms of the androgen receptor pathway in therapy-resistant prostate cancer cell models. *PLoS One* 5: e13500.
59. Debes JD, Tindall DJ (2004) Mechanisms of androgen-refractory prostate cancer. *N Engl J Med* 351: 1488–1490.
60. Feldman BJ, Feldman D (2001) The development of androgen-independent prostate cancer. *Nat Rev Cancer* 1: 34–45.
61. Heinlein CA, Chang C (2004) Androgen receptor in prostate cancer. *Endocr Rev* 25: 276–308.
62. Ueda T, Mawji NR, Bruchovsky N, Sadar MD (2002) Ligand-independent activation of the androgen receptor by interleukin-6 and the role of steroid receptor coactivator-1 in prostate cancer cells. *J Biol Chem* 277: 38087–38094.
63. Baca SC, Garraway LA (2012) The genomic landscape of prostate cancer. *Front Endocrinol (Lausanne)* 3: 69.
64. Wang Q, Li W, Zhang Y, Yuan X, Xu K, et al. (2009) Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* 138: 245–256.
65. Lorient Y, Wyatt A, Al Nakouzi N, Beraldi E, Toren P, et al. (2013) Mechanisms of resistance to enzalutamide in LNCaP models. *Cancer Res* 73.
66. Massie CE, Adryan B, Barbosa-Morais NL, Lynch AG, Tran MG, et al. (2007) New androgen receptor genomic targets show an interaction with the ETS1 transcription factor. *EMBO Rep* 8: 871–878.
67. Chen H, Libertini SJ, George M, Dandekar S, Tepper CG, et al. (2010) Genome-wide analysis of androgen receptor binding and gene regulation in two CWR22-derived prostate cancer cell lines. *Endocr Relat Cancer* 17: 857–873.
68. Takayama K, Kaneshiro K, Tsutsumi S, Horie-Inoue K, Ikeda K, et al. (2007) Identification of novel androgen response genes in prostate cancer cells by coupling chromatin immunoprecipitation and genomic microarray analysis. *Oncogene* 26: 4453–4463.
69. Takayama K, Tsutsumi S, Katayama S, Okayama T, Horie-Inoue K, et al. (2011) Integration of cap analysis of gene expression and chromatin immunoprecipitation analysis on array reveals genome-wide androgen receptor signaling in prostate cancer cells. *Oncogene* 30: 619–630.
70. van Brabant AJ, Stan R, Ellis NA (2000) DNA helicases, genomic instability, and human genetic disease. *Annu Rev Genomics Hum Genet* 1: 409–459.
71. Chymkowitch P, Le May N, Charneau P, Compe E, Egly JM (2011) The phosphorylation of the androgen receptor by TFIIH directs the ubiquitin/proteasome process. *The EMBO journal* 30: 468–479.
72. Hubert L Jr, Lin Y, Dion V, Wilson JH (2011) Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1. *Hum Mol Genet* 20: 4822–4830.
73. Lin Y, Wilson JH (2009) Diverse effects of individual mismatch repair components on transcription-induced CAG repeat instability in human cells. *DNA repair* 8: 878–885.
74. Mu XM, Chang CS (2003) TR2 orphan receptor functions as negative modulator for androgen receptor in prostate cancer cells PC-3. *Prostate* 57: 129–133.
75. Russo D, Durante C, Bulotta S, Puppini C, Puxeddu E, et al. (2013) Targeting histone deacetylase in thyroid cancer. Expert opinion on therapeutic targets 17: 179–193.
76. Palvimo JJ, Reinikainen P, Ikonen T, Kallio RJ, Moilanen A, et al. (1996) Mutual transcriptional interference between RelA and androgen receptor. *Journal of Biological Chemistry* 271: 24151–24156.
77. de Moissac D, Zheng H, Kirshenbaum LA (1999) Linkage of the BH4 domain of Bcl-2 and the nuclear factor kappaB signaling pathway for suppression of apoptosis. *J Biol Chem* 274: 29505–29509.
78. Abern MR, Bassett MR, Tsvivan M, Banez LL, Polascik TJ, et al. (2013) Race is associated with discontinuation of active surveillance of low-risk prostate cancer: Results from the Duke Prostate Center. *Prostate Cancer Prostatic Dis* 16: 85–90.

79. Fukagai T, Namiki T, Carlile RG, Namiki M (2009) Racial differences in clinical outcome after prostate cancer treatment. *Methods in molecular biology* 472: 455–466.
80. Powell IJ (1998) Prostate cancer in the African American: is this a different disease? *Seminars in urologic oncology* 16: 221–226.
81. Chornokur G, Dalton K, Borysova ME, Kumar NB (2011) Disparities at presentation, diagnosis, treatment, and survival in African American men, affected by prostate cancer. *The Prostate* 71: 985–997.
82. Ellis SD, Blackard B, Carpenter WR, Mishel M, Chen RC, et al. (2013) Receipt of National Comprehensive Cancer Network guideline-concordant prostate cancer care among African American and Caucasian American men in North Carolina. *Cancer* 119: 2282–2290.
83. Kubricht WS 3rd, Williams BJ, Whatley T, Pinckard P, Eastham JA (1999) Serum testosterone levels in African-American and white men undergoing prostate biopsy. *Urology* 54: 1035–1038.
84. Marks LS, Hess DL, Dorey FJ, Macairan ML (2006) Prostatic tissue testosterone and dihydrotestosterone in African-American and white men. *Urology* 68: 337–341.
85. Gaston KE, Kim D, Singh S, Ford OH 3rd, Mohler JL (2003) Racial differences in androgen receptor protein expression in men with clinically localized prostate cancer. *J Urol* 170: 990–993.
86. Kim HS, Moreira DM, Jayachandran J, Gerber L, Banez LL, et al. (2011) Prostate biopsies from black men express higher levels of aggressive disease biomarkers than prostate biopsies from white men. *Prostate Cancer Prostatic Dis* 14: 262–265.
87. Lange EM, Sarma AV, Ray A, Wang Y, Ho LA, et al. (2008) The androgen receptor CAG and GGN repeat polymorphisms and prostate cancer susceptibility in African-American men: results from the Flint Men's Health Study. *Journal of human genetics* 53: 220–226.
88. Bratt O, Borg A, Kristofferson U, Lundgren R, Zhang QX, et al. (1999) CAG repeat length in the androgen receptor gene is related to age at diagnosis of prostate cancer and response to endocrine therapy, but not to prostate cancer risk. *British Journal of Cancer* 81: 672–676.
89. Ding D, Xu L, Menon M, Reddy GP, Barrack ER (2004) Effect of a short CAG (glutamine) repeat on human androgen receptor function. *The Prostate* 58: 23–32.
90. Rodriguez-Gonzalez G, Cabrera S, Ramirez-Moreno R, Bilbao C, Diaz-Chico JC, et al. (2009) Short alleles of both GGN and CAG repeats at the exon-1 of the androgen receptor gene are associated to increased PSA staining and a higher Gleason score in human prostatic cancer. *J Steroid Biochem Mol Biol* 113: 85–91.