AD_____

Award Number: W81XWH-11-1-0555

TITLE: Understanding the Collaboration between the Androgen Receptor and ERG in Prostate Cancer

PRINCIPAL INVESTIGATOR: Alok K. Tewari

CONTRACTING ORGANIZATION:
Duke University Medical Center

Durham, NC 27705

REPORT DATE: August 2012

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
August 2012	Annual Summary	15 July 2011- 14 July 2012

4. TITLE AND SUBTITLE

Understanding the Collaboration between the Androgen Receptor and ERG in Prostate Cancer

5a. CONTRACT NUMBER
W81XWH-11-1-0555

5b. GRANT NUMBER
W81XWH-11-1-0555

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Alok K Tewari

E-Mail: alok.tewari@duke.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Duke University Medical Center

Durham, NC 27705

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Epigenetic mechanisms such as chromatin accessibility and histone modifications impact transcription factor binding to DNA and transcriptional specificity. The androgen receptor (AR), a master regulator of the male phenotype and prostate cancer pathogenesis, acts primarily through ligand-activated transcription of target genes. Our purpose is to explore how the ERG transcription factor impacts AR function by focusing on chromatin accessibility, AR binding and transcription. The first year of this award has been spent focusing on developing and optimizing the molecular biology techniques, computational tools and analyses to understand the chromatin dynamics of AR activation. This work lays a critical groundwork for understanding how perturbations to prostate cancer cells, such as ERG over-expression, impacts chromatin structure, AR binding to DNA, and transcription. Our analysis of AR activation in prostate cancer cells revealed both qualitative and quantitative differences in chromatin accessibility that largely correspond to AR binding and AR-regulated RNA transcription across the genome and underscores the importance of analyzing chromatin data as continuous signal rather than discrete binary peaks. Intriguingly, base pair resolution of the DNaseI cleavage profile revealed distinct footprinting patterns associated with the AR-DNA interaction, suggesting multiple modes of AR-DNA interactions. These analyses suggest a dynamic and novel quantitative relationship between chromatin structure and AR-DNA binding which impacts AR transcriptional specificity. The results from this year of work have been presented at national meetings are under review for publication.

15. SUBJECT TERMS
Androgen Receptor, DNaseI hypersensitivity, chromatin structure, prostate cancer

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	21	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Introduction

The androgen receptor (AR), a member of the nuclear receptor superfamily, is a master regulator of prostate cancer. Prostate cancer (PC) is the most common non-cutaneous neoplasm affecting males in the United States [1]. Due to the dependence of PC on androgen, primary therapy for non-localized and recurrent PC is aimed at chemically abrogating androgen signaling through its receptor, the AR, either by reducing the amount of androgen ligand available to activate AR signaling, or directly antagonizing the AR. These strategies inhibit disease progression for a variable period of time, but progression inevitably occurs, often through continued AR-mediated signaling**.** Indeed, several lines of evidence suggest that both localized and castration-resistant prostate cancer (CRPC) – which is generally defined as disease that has progressed despite standard androgen deprivation therapies [2] - are critically dependent on continued AR activity [3-7]. Several mechanisms contribute to continued AR signaling in advanced disease, including *de novo* synthesis of androgen [8], copy number gain [9], point mutations [9-11], alterations to the dynamic balance of AR co-activators and co-repressors through a variety of mechanisms [9, 12-14], and crosstalk with various kinase signaling pathways [15, 16]. As a result, there is great enthusiasm over the recent development of novel and more effective therapies that target AR activity in CRPC such as abiraterone acetate and enzalutamide. As the AR acts primarily through transcriptional regulation of target genes, it is critical to further understand the determinants of the AR-mediated transcriptional program.

Since the first discovery of fusion of the 5' regulatory region of the androgen-regulated TMPRSS2 gene to the ETS family members ERG (Ets-related gene) and ETV1 (Ets-variant1), significant strides have been made in understanding the prevalence and characteristics of these fusions, summarized in [17]. Large cohort studies suggest that approximately 50% of prostate cancers contain recurrent TMPRSS2-ERG gene fusions, generally characterized by 5' genomic elements that are either expressed at high levels under the control of androgen, fused to portions of ETS family members. These fusions lead to the overexpression of ETS family members, with TMPRSS2-ERG the most common gene fusion product identified. The prognostic significance of these fusions is unclear, although the largest study to date of men in the United States treated with surgery suggests no relationship between fusion status and more aggressive prostate cancer [18].

Although ERG alone exerts a modest impact on prostate epithelial cells, numerous studies have demonstrated collaboration between AR and ERG expression in prostate cancer. Multiple studies in mice demonstrated that co-overexpression of both genes synergized to generate invasive prostate cancer [19, 20]. Convincing evidence exists to suggest a role for chromatin structure in the cooperation between the AR and ERG. *In silico* analysis of ERG co-expression patterns in a cohort of tumors revealed that HDAC1 is consistently highly expressed along with ERG [21]. The importance of histone deacetylase proteins in the ERG fusion-positive tumor context was confirmed by the same group in a later study showing that HDAC inhibitory compounds are effective in slowing the growth of ERG-fusion positive cell lines *in vitro* [22]. Furthermore, two studies recently showed that AR and ERG co-occupy thousands of locations across the genome, and many of these locations are associated with AR-regulated genes [23, 24]. Both studies propose a model whereby ERG expression serves to repress and thus re-program AR function to cause prostate cell de-differentiation, putatively promoting prostate tumorigenesis. The more recent of these studies further found that various HDACs including HDAC1 co-occupy genomic regions bound by the AR and ERG. Finally, two reports suggest that AR function may prime prostate epithelial cells to be predisposed to generation of ERG fusions [25, 26]. These findings suggest that the temporal relationship in prostate cancer progression is first differentiation of basal epithelial cells towards luminal epithelial cells such that they acquire the AR, followed by generation of ERG fusion products. This putative

mechanism is reinforced by the relatively low prevalence of ERG fusions in high-grade PIN compared to local adenocarcinoma [27, 28]. Together, these various lines of evidence point to a collaborative role between the AR, ERG and histone deacetylates proteins in prostate cancer. However, these studies have focused only on cell lines derived from metastatic tumors, and do not fully explore this cooperation in prostate cancer initiation.

DNaseI hypersensitivity (DHS) analysis, based on the preferential cleavage of euchromatin by the DNase I enzyme, offers a tool to interrogate chromatin structure that has been recognized as a marker for accessible chromatin such as that seen at promoters, enhancers, silencers and locus control regions [29]. Coupling this assay with high throughput sequencing technologies (DNase-seq) offers the ability to interrogate chromatin structure on a genome-wide scale [30, 31]. Such studies have further found that DNaseI HS correlates with transcriptionally relevant histone modifications such as H3K4me2 [30] and can also be used to identify DNaseI footprints that correlate to transcription factor binding [32]. DNaseI HS analysis has been used to detect chromatin structure around the TSS of various genes impacted by the ligand of different nuclear receptors, including those in the same family as the AR. For example, DNase-seq analysis of glucocorticoid receptor (GR) activation revealed that the GR initially targets regions of chromatin that are accessible prior to activation. A majority of these pre-accessible regions of chromatin are poised by occupancy of the AP-1 protein [33, 34]

Given the evidence supporting the interplay between AR and ERG in prostate cancer at the level of chromatin and the proven ability of high-throughput chromatin assays to uncover relevant biology, this research proposal aims to integrate chromatin accessibility, protein-DNA binding and transcription to further elucidate how ERG impacts AR function in prostate cancer initiation.

Body

I. Optimization of techniques and development of computational analyses

In order to move forward with the specific aims outlines in the Statement of Work, we first had to develop an analytical workflow to effectively integrate DNase-seq, ChIP-seq and mRNA-seq data. This effort has been the focus of the first year of work.

To develop such a workflow, we focused on LNCaP cells before (LNCaP) and after (LNCaP-induced) AR activation by 1 nM of the synthetic ligand R1881 for 12 hours. LNCaP cells grow easily in cell culture, and are a canonical model for AR activation. We generated a more deeply sequenced data set including more biological and technical replicates. These replicates serve to make our data set more robust and reproducible, and enable more detailed analysis. A summary of the data generated is shown below in Table 1.

	LNCaP	**LNCaP Induced**
Total DNase-seq Reads	129,131,592	138,464,636
Number of DHS	144,070	140,966
Bases within DHS	86,989,168	82,887,882
Percentage of genome	3.01	2.87

Table 1: Summary of DNase-seq experiments. Three biological replicates of LNCaP and two biological replicates of LNCaP-induced were combined to create final DNase-seq libraries.

Inspection of our data revealed that chromatin accessibility is more nuanced than a simple open or closed state. Thus, we approached interpretation of DNase-seq data in two ways: (1) calling discrete peaks, referred to as DNaseI hypersensitive (DHS) sites, and

comparing regions qualitatively as binary conditions (DHS site or not), and (2) identifying regions of statistically different DNase-seq signal before and after hormone treatment, referred to as ΔDNase regions. For the first method, we used our previously established analytical pipeline to identify DHS sites [35]. For the second method, we chose to utilize the edgeR algorithm [36] to detect significant changes in signal. These analyses generated regions that we could relate to AR binding and transcription data.

To generate sequence-based transcriptome data for AR activation, we harvested mRNA from multiple biological replicates under the same growth conditions as the DNase-seq data and created sequencing libraries using the standard mRNA-seq protocol provided by Illumina. To analyze the resultant sequencing data, we integrated multiple available bioinformatics tools into a single pipeline, as illustrated below in Figure 1.

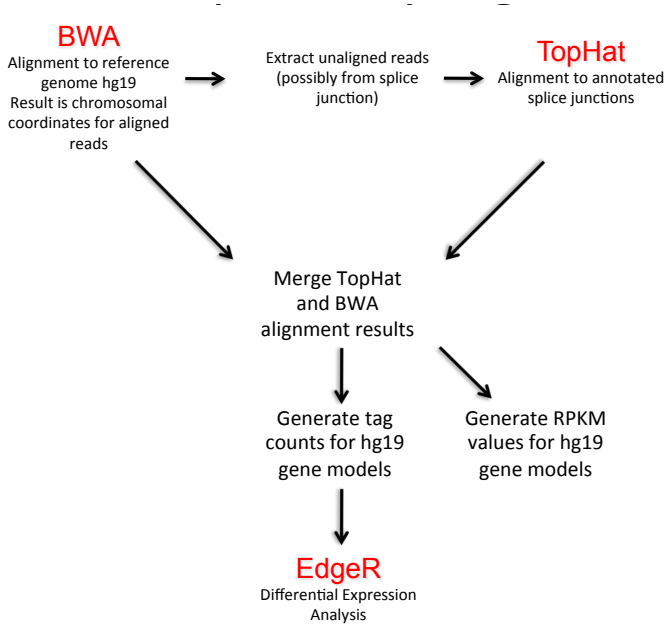

Figure 1: mRNA-seq analysis pipeline. Raw sequence reads were aligned to the reference genome using BWA [37] and unaligned reads were further matched to splice junctions using TopHat [38]. Mapped reads were matched to established gene models, and expression values were generated in the RPKM (reads per kilobase mapped) scale. Differential expression analysis was carried out with edgeR.

Finally, we spent considerable time attempting to establish ChIP-seq. While we were able to generate fixed chromatin, we were unable to adequately fragment our chromatin down to the 500 bp range required to create the sequencing library despite extensive troubleshooting. This is an ongoing effort. Fortunately, we were able to utilize three published AR ChIP-seq data sets generated under various growth conditions as shown in Table 2. We therefore utilized three sets of AR ChIP-seq data from LNCaP cells (Table 2) that we refer to as "Yu" [24], "Massie" [39] and "Coetzee" [40, 41]. To minimize the impact of technical variation within each individual experiment, we created two high confidence sets of AR binding sites from these three sources: (i) an “R1881 intersect” set consisting of Yu and Massie peaks that overlap each other, as these experiments used the same AR hormone ligand as our DNase-seq experiments (R1881); and (ii) an “All AR intersect” data set containing the intersection of peaks from all three data sets including the Coetzee experiment that used an alternative AR ligand, dihydrotestosterone (DHT).

Data Set	Ligand	Treatment time	No. of AR binding sites
Massie	1 nM R1881	4 hr	19,505
Yu	10 nM R1881	16 hr	37,676
Coetzee	10 nM DHT	4 hr	12,929
R1881 intersect (Massie/Yu intersect)	R1881		13,258
All AR Intersect	R1881/DHT		5,940

Table 2: Characteristics of AR ChIP-seq data sets. Name, ligand, ligand treatment time and number of peaks called for each AR ChIP-seq data set are shown. R1881 intersect represents the intersection of the Massie and Yu data sets. All AR intersect represents high confidence AR binding sites that are found in all three data sets.

Through these efforts, we were able to generate new analysis pipelines for DNase-seq, mRNA-seq and ChIP-seq data and integrate these data sets into an exploration of how AR activation impacts chromatin accessibility, AR binding, and transcription.

II. Results from analyses to date

DNase-seq identifies changes in chromatin accessibility with androgen receptor activation

From our robust DNase-seq data we identified 144,070 DHS sites in LNCaP and 140,966 DHS in LNCaP-induced cells using a p-value cutoff of 0.05 (Table 1). A comparison of the DHS sites identified in LNCaP-induced and LNCaP reveals that 102,173 (72.5%) overlap. To put the degree of overlap in context, we used the same criteria to identify DHS sites in 7 unrelated cell lines for which high quality DNase-seq data is available (NHEK, GM12678, HelaS3, HepG2, HUVEC, K562, and H1-ES) [42]. The average overlap between distinct cell lines is 50.4% +/- 7.04%, which is substantially less than the overlap between LNCaP and LNCaP-induced cells (Figures 2A and 2B). We also investigated the overall distribution of DHS sites relative to promoters, intronic, or intergenic regions and found that the location of all DHS sites prior to and following AR activation does not shift this distribution. These data suggest that while AR activation induces a modest amount of chromatin changes, the overall degree of these changes are substantially less than those detected between cell lines from unrelated tissues.

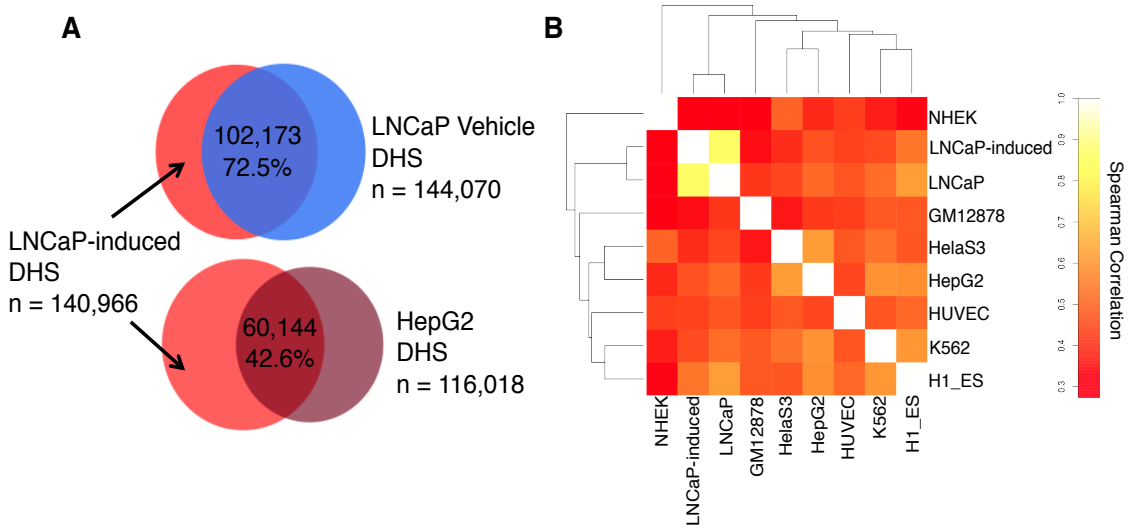

Figure 2: Changes in chromatin accessibility with AR-activation. (A) Overlap between DHS sites identified before (Vehicle or LNCaP) and after hormone (LNCaP-induced) as compared to the unrelated liver carcinoma cell line HepG2. (B) Spearman correlation heatmap of DNase-seq score in the union set of top 100,000 DHS peaks in each of the 9 cell lines illustrated.

As discussed above, in order to quantitatively identify those loci with the most substantive increase or decrease in DNase-seq signal with AR activation, we used the edgeR statistical package. Increases represent regions that become more accessible after hormone treatment, and decreases become less accessible. To capture a broad spectrum of significant changes in signal, we used two statistical thresholds ("strict" = a false discovery rate (FDR) threshold of 5%, and "loose" = unadjusted p-value threshold of 0.05) to identify the degree of accessibility changes, which we refer to as ΔDNase regions (Table 3). These regions suggest that AR activation results primarily in regions with increased rather than decreased chromatin accessibility. Mapping all regions of significantly changed DNase-seq signal to genic elements revealed a depletion of promoter regions and enrichment for both inter- and intra-genic locations compared to all DHS sites with AR activation (Figure 3A). An example of one such ΔDNase increase is shown below in Figure 3B at the well-described *KLK3* AR-binding enhancer element.

Strict Threshold	**Number of regions**
Strict ΔDNase increase	2,586
Strict ΔDNase decrease	0
Loose Threshold	
Loose ΔDNase increase	18,692
Loose ΔDNase decrease	1,467

Table 3: Number of Differential regions of DNase-seq with AR activation (ΔDNase). ΔDNase regions were identified using edgeR. Strict ΔDNase increases are a complete subset of the loose ΔDNase increase regions.

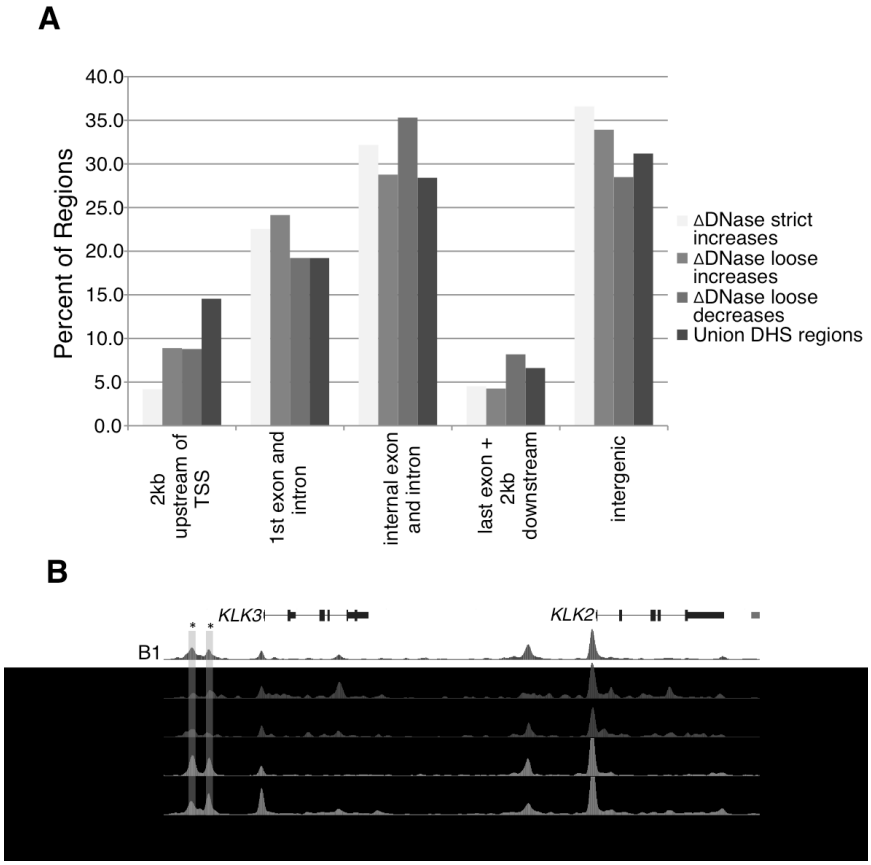

Figure 3: (A) Distribution of ΔDNase regions and union (LNCaP and LNCaP-induced) DHS sites relative to genic elements. (B) Replicates of DNase-seq data around *KLK3* and *KLK2*. Y-axis is fixed for all rows. Highlighted regions marked by an asterisk represent examples of significant ΔDNase increases.

We hypothesized that ΔDNase regions represented locations where AR activation altered transcription factor binding. As expected, we found a strong AR motif match in regions of increased chromatin accessibility. In addition, several other significantly enriched motifs were detected in both ΔDNase increase and decrease regions (Figure 4) that correspond to transcription factors such as SP1. SP1 can bind directly with multiple known AR co-factors as well as the AR [43] and represents an intriguing protein for further investigation as a modifier of AR function.

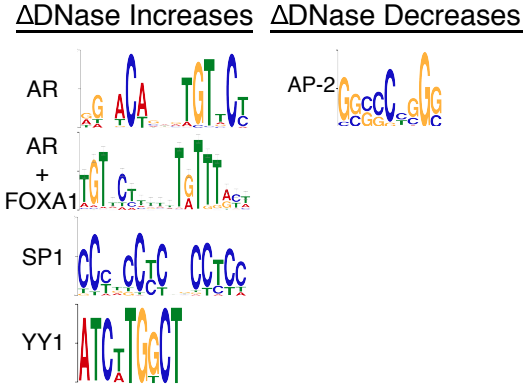

Figure 4: *De novo* motif analysis identifies putative transcription factors that may impact AR-induced chromatin accessibility and thus AR function.

The androgen receptor binds both poised and remodeled chromatin accessible to DNaseI cleavage

We next examined the relationship between chromatin accessibility and AR binding to the genome. Each of the three individual AR ChIP studies displayed consistent overlap patterns with DHS sites. In each individual experiment approximately 20% of all AR binding sites occurred within DHS sites that are present both before and after hormone treatment ("poised") and an additional 20-30% of AR binding sites overlapped DHS sites following androgen induction. Thus, each data set suggests that slightly less than half of all AR binding sites in DHS regions are poised (Figures 5A, 5B) and the remainder change in response to androgen treatment. The high confidence AR (R1881 intersect and All AR intersect) binding sites displayed a similar trend. Of note, only 1-2% of AR binding sites map within a DHS site present in LNCaP, but not LNCaP-induced cells.

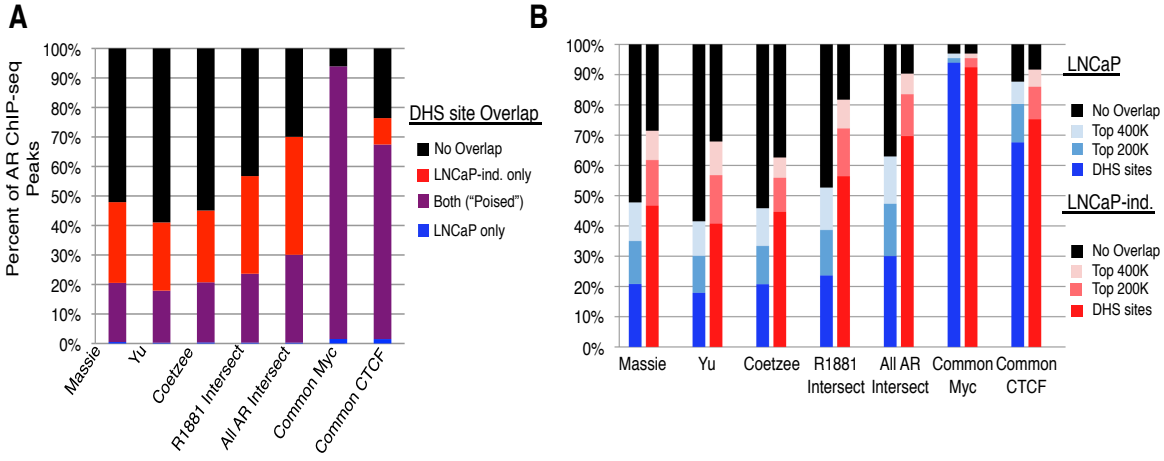

Figure 5: Relationship between AR binding and DHS regions. (A) Overlap of each ChIP-seq data set with poised LNCaP DHS (regions that are DHS sites in both LNCaP and LNCaP-induced, shown in purple) and LNCaP-induced only DHS sites (shown in red). AR binding sites not overlapping a DHS site are represented in black. Common Myc and CTCF binding sites are shown as controls. (B) Overlap of ChIP-seq peaks shown at different thresholds of DNase-seq enrichment ("DHS sites" representing the regions of significant signal over background $p<0.05$, "Top 200k" representing the top 200,000 initial peaks showing enrichment over background, and "Top 400k" representing all regions showing DNase-seq enrichment over background). Columns in various shades of blue show overlap with LNCaP DHS at different thresholds, and columns in various shades of red show overlap with LNCaP-induced DHS at different thresholds.

The amount of AR binding to both poised and LNCaP-induced DHS sites is in stark contrast to Myc and CTCF binding sites [42] that almost exclusively bind within poised DHS sites (Figure 5A). Thus, of the AR binding events occurring within a DHS site, less than half occurred in poised regions, with the majority binding to regions that displayed qualitative AR induced chromatin remodeling.

Given the observation that a substantial number of AR binding sites occur within LNCaP-induced only DHS sites, we examined the association between AR binding events and quantitative chromatin remodeling. To test this, we evaluated AR sites that overlapped regions with increased DNase-seq signal (strict and loose ΔDNase increases). As expected, AR ChIP-

seq peaks identified only within LNCaP-induced DHS sites (Circle III, Figure 6A) show significant overlap with ΔDNase increase regions. Interestingly, AR binding sites in peaks found in both LNCaP and LNCaP-induced cells (Circle II, Figure 6A) were also enriched for ΔDNase increases, although not to the same extent as those sites that mapped only within LNCaP-induced DHS. The proportion of AR binding regions that mapped to poised, LNCaP-induced DHS sites only, and ΔDNase regions were consistent across each AR binding data set. Analogously, we found that 36.5% of strict ΔDNase increases and 16.7% of loose ΔDNase increases overlapped the high confidence AR binding set ("All AR intersect") (Figure 6B). These observations indicate that even if AR binding occurred within poised chromatin, these binding events were associated with a substantial increase in chromatin accessibility, highlighting the utility of identifying regions of ΔDNase signal. These findings support similar observations at three previously identified "poised" AR enhancers [40] and suggest that AR binding more globally promotes chromatin accessibility, allowing for more DNaseI cleavage following hormone treatment.

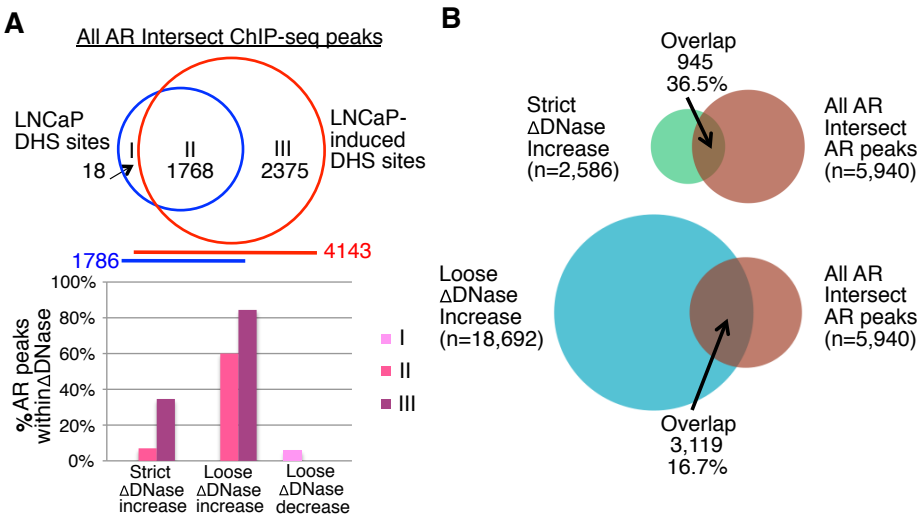

Figure 6: Overlap of AR binding sites and ΔDNase regions. (A) Venn diagram shows overlap of DHS sites and the high confidence "All AR Intersect" data set. Column plots below illustrate the overlap of each region of the Venn diagram (I – AR binding sites only in an uninduced LNCaP DHS site, II – AR binding sites that fall within both a LNCaP and LNCaP-induced DHS sites (poised), III – AR binding sites in LNCaP-induced only DHS sites) with ΔDNase regions. (B) The reverse comparison illustrating the overlap of ΔDNase regions with AR binding sites.

Data from Figure 5 indicates that a proportion of AR binding occurs in non-DHS sites, and that a small yet significant subset of AR binding occurs in regions of the genome inaccessible to cleavage by the DNaseI enzyme. We thus wondered whether there was a difference in the quality of AR binding within DHS sites relative to less accessible chromatin. Indeed, AR binding signal was stronger in regions overlapping DHS sites than non-DHS regions (Figure 7), and was the strongest for AR sites common to two or three experiments. Motif analysis of regions that bound the AR but were inaccessible to DNaseI cleavage at any DNase-seq threshold revealed a very similar binding motif to the canonical AR DNA recognition sequence. Thus, it appears that AR binding occurs at a range of chromatin accessibility and accessibility correlates with AR binding strength.

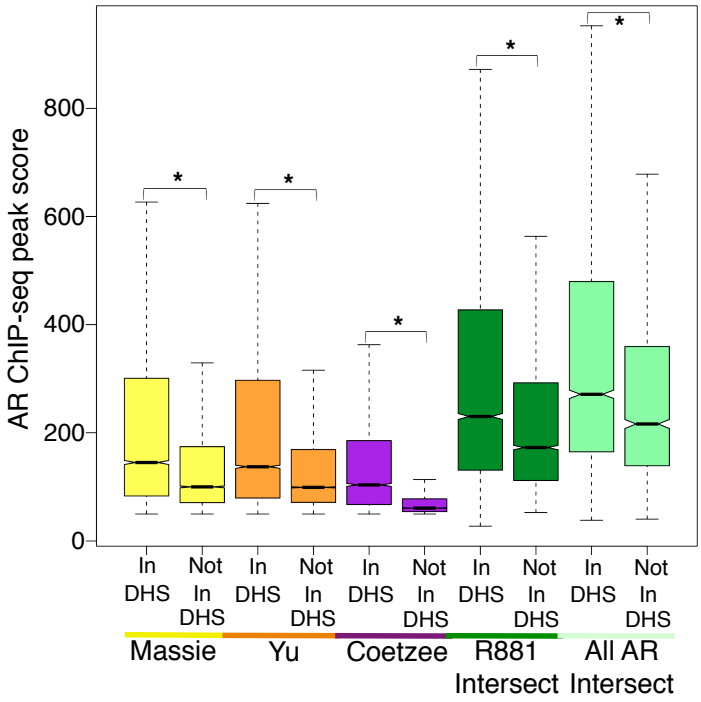

Figure 7: AR ChIP-seq binding scores for peaks overlapping and not overlapping DHS sites. Asterisks denote significant differences in AR peak score (Mann-Whitney p-value < 0.001)

Changes in chromatin accessibility correlate with the AR transcriptional program.

To determine if ΔDNase regions were associated with AR-mediated transcriptional changes, we analyzed our mRNA-seq data and identified genes differentially regulated by androgen induction. Expression values from three replicates generated clustered according to hormone treatment status (Figure 8A). Using edgeR [36], we identified 339 genes differentially expressed upon AR induction (FDR < 0.05), 202 of which were upregulated and 137 of which were downregulated (Figure 8B). Of these, 46% were identified as AR target genes in at least one other published microarray study.

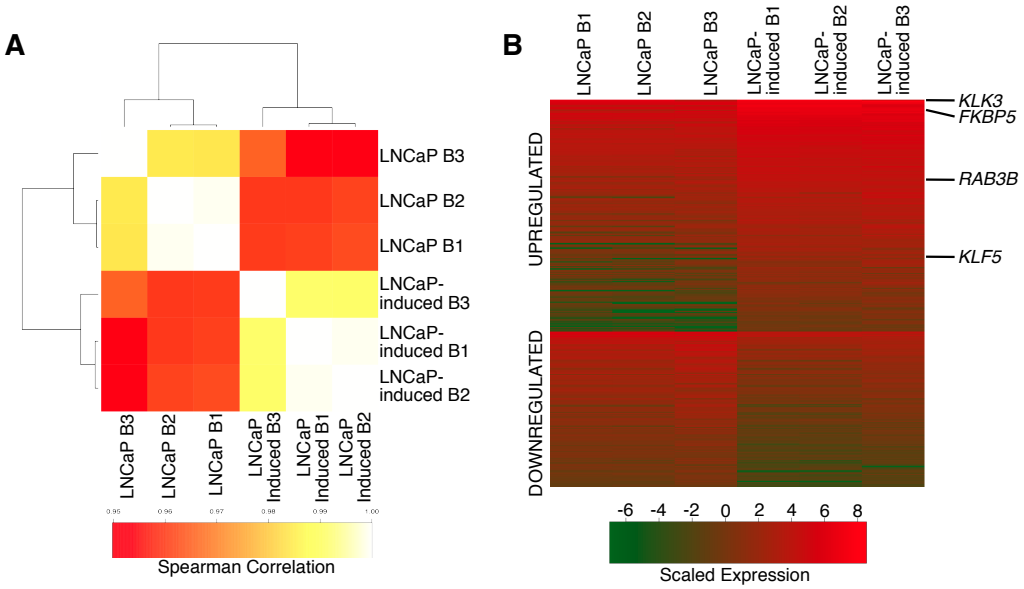

Figure 8: mRNA-seq analysis of AR-mediated transcriptional changes. (A) Spearman correlation heatmap of expression (RPKM) data from each biological replicate of LNCaP and LNCaP-induced cells. (B) Heatmap of expression levels (RPKM) for genes identified as differentially regulated by the AR. Rows are ordered by total sum. Genes most commonly identified in microarray studies as AR-regulated are all located near the top of the heatmap, reflecting their overall high levels of expression before and after androgen induction.

We hypothesized that AR-mediated changes in chromatin accessibility contribute to the AR-mediated gene expression program. By mapping ΔDNase regions to their closest gene, we found that strict ΔDNase increase regions were significantly enriched near up-regulated genes (p-value < 0.001) and were modestly enriched with downregulated genes (p = 0.053; Figure 9A). Loose ΔDNase increases were significantly enriched near both up- and down-regulated genes (p < 0.001). Loose ΔDNase decreases were not present near upregulated genes, but were modestly enriched with downregulated genes (p-value = 0.057). We performed an identical analysis using ΔDNase regions and microarray expression data from Massie et al. [39], and observed similar associations (Figure 9B). The reverse comparison wherein we associated differentially regulated genes to ΔDNase regions within 20 kb of the transcriptional start site shows a similar trend (Figures 9C, 9D). Both up- and downregulated genes were associated with loose ΔDNase increases in chromatin accessibility. Interestingly, the borderline significant associations between AR downregulated genes and strict ΔDNase increases as well as loose ΔDNase decreases became very insignificant upon limiting the distance criteria for associating a ΔDNase region to a gene (Figure 9C). This finding may indicate that AR-mediated repression of gene expression requires chromatin interactions over longer distances. Overall, our data support the hypothesis that AR activation preferentially causes distal chromatin accessibility changes that are significantly associated with nearby gene expression changes.

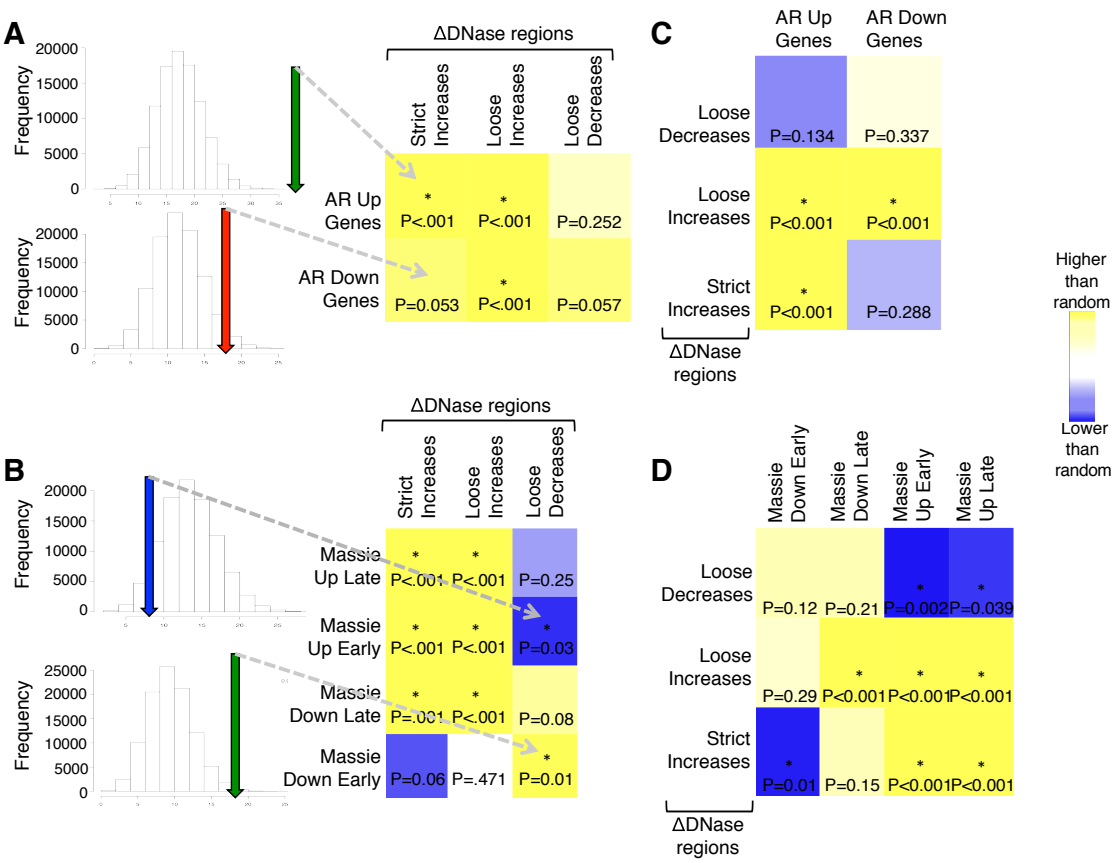

Figure 9: Association between ΔDNase regions and AR-regulated transcription. (A-B) ΔDNase regions were mapped to the closest gene, and the amount of overlap between these genes and a randomly chosen gene set containing the same number of genes as were identified as AR-regulated was permuted 100,000 times to generate a null distribution (histograms) and assess significance of the overlap between ΔDNase associated genes and AR-regulated genes. Arrows indicate the actual overlap between ΔDNase nearest genes and AR-regulated genes from either mRNA-seq analysis or Massie et al. Blue shading represents less ΔDNase regions (absence/depletion) around AR-regulated genes, whereas yellow shading represents more ΔDNase regions (presence/enrichment) around AR-regulated genes. (C-D) The reverse comparison, relating AR-regulated expression to ΔDNase regions for mRNA-seq and Massie et al. data.

Base-pair resolution analysis of DNase-seq reveals multiple signal profiles

Our group and others have shown that deep sequencing of DNaseI cleavage libraries can detect individual transcription factor binding events via the identification of DNaseI footprints and that DNaseI footprints correspond to local protection of DNA from nuclease cleavage by bound transcription factors [32, 44, 45]. To examine the AR-DNA binding footprint, we examined the aggregate DNase-seq signal around the AR-DNA recognition motif within AR binding sites and compared the resultant pattern to that around the recognition motif of other transcription factors within our data. An overall increase in DNase signal was observed around AR motifs (Figure 10A) compared to other transcription factor motifs such as CTCF and NRSF (Figures 10B and 10C). A symmetrical depletion of DNase-seq signal was detected around AR motifs in DHS sites that closely matches the information content of the AR binding motif dimer (Figure

10A, red line) [46]. In poised AR binding sites, we observed a similar pattern of protection despite lower overall DNase-seq signal intensity (Figure 10A, blue line). Binding sites that became available only after androgen induction only exhibited the footprint after androgen treatment (Figure 10D, blue line). Importantly, the overall enrichment of DNase signal in LNCaP-induced cells is specific to DHS regions that bind the AR and have an AR motif, as opposed to all DHS sites (Figure 10E).

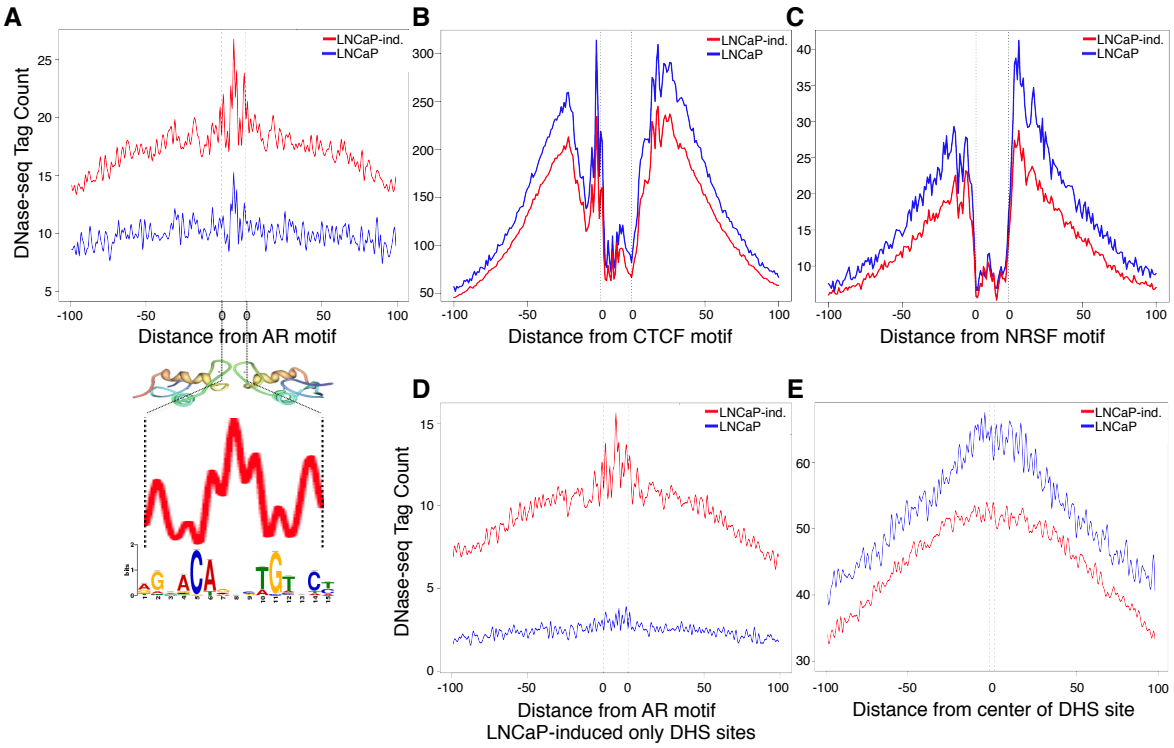

Figure 10: Base pair resolution around AR motif matches reveals a unique pattern of protection by the AR. (A) Aggregate DNase-seq signal around AR motif matches within poised DHS sites that bind the AR. The pattern of DNaseI cuts within the motif closely follows the known structure of the AR dimer as well as the information content of the AR DNA recognition motif determined from ChIP-seq data. Aggregate DNase-seq signal centered on CTCF (B) and NRSF (C) motif matches genome-wide display a structurally different footprint from that of the AR. (D) Aggregate signal around AR motif matches within DHS sites unique to LNCaP-induced cells that also bind the AR. (E) Aggregate signal around the center of 10,000 randomly chosen DHS sites shared between LNCaP and LNCaP-induced cells. Note that overall the aggregate signal is higher in LNCaP as compared to LNCaP-induced cells within all DHS sites.

Two algorithms exist to identify transcription factor binding events from DNase-seq data. These algorithms are able to successfully predict binding of large transcription factors with a strong footprint (such as CTCF and NRSF – see Figure 10). We attempted to apply these algorithms to prospectively identify AR binding events with minimal success. Reasoning that since these algorithms learn the footprint pattern from the aggregate DNase-seq signal at a motif, such as shown in Figure 10A, we supposed that discrete patterns of AR binding and subsequent protection from DNaseI cleavage may occur at specific loci, thus confounding the models. Indeed, using k-means clustering we identified three reproducible clusters of DNaseI

cleavage patterns, each of which represented part of the observed composite footprint (Figure 11). These clusters were much less frequently detected across repeated iterations of clustering in untreated LNCaP cells, indicating the three distinct patterns of DNaseI protection appeared to be a robust phenomenon more often detected in LNCaP-induced DNase-seq data, suggesting that AR activation stabilizes specific chromatin structure around AR motifs. AR binding has been associated with enrichment of palindromic full-site AR motifs (such as depicted in Figure 10A) as well as half-site motifs [47, 48]. The directional footprinting in clusters 1 and 2 is indicative of only half of the full canonical AR motif being protected from DNaseI cleavage, whereas Cluster 3 is consistent with full-site protection. Our ability to detect this indicates that specific half site usage is consistent across the entire population of cells, and does not fluctuate randomly. The spike in the center of Cluster 3 corresponds to the degenerate bases in the middle of the AR motif, indicating reduced DNA protection between AR proteins, possibly within a dimer. Overall, our footprinting analysis revealed three different stable modes of AR binding that represent either full or half-site protection at full-site DNA motifs.

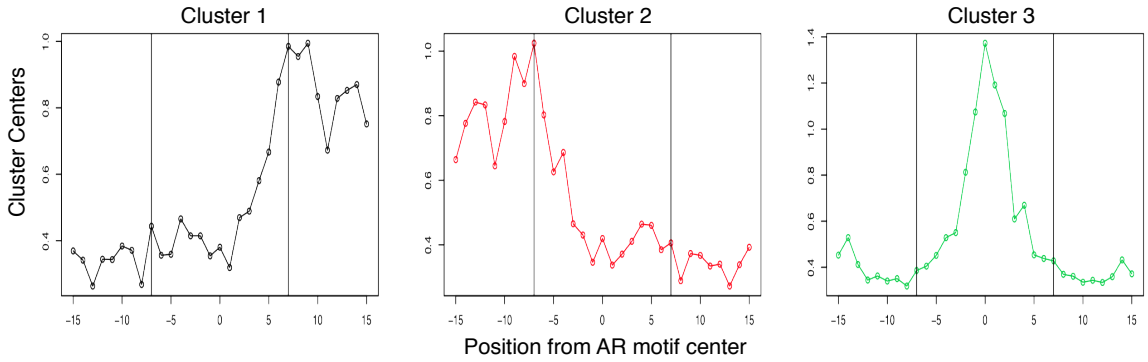

Figure 11: K-means clustering of LNCaP-induced DNase-seq signal into three consistent clusters within AR binding sites.

III. Generation of DNase-seq and expression data for PrEC +/- AR activation

To utilize the computational framework developed towards understanding the role of AR and ERG in prostate cancer tumorigenesis, we proceeded with Aim 2 from the Project Narrative utilizing an immortalize and tumorigenic prostate epithelial cell line expressing the AR. From previous work in our lab, we knew that these cells differentiate and decrease their rate of proliferation with AR activation by ligand [49]. This experiment was to serve as proof of principle that these cell lines, which are more difficult to handle than other established prostate cancer cell line models, could be processed for DNase-seq analysis. Preparation of DNase-seq libraries was successful and genome-wide data was generated. However, in cross-replicate analysis and subsequent analysis of AR-mediated transcription, it became clear that AR activation was not inducing known transcriptional changes. We are currently troubleshooting this approach to ensure adequate AR activation before moving forward.

Key Research Accomplishments

- Established molecular biology techniques for DNase-seq and mRNA-seq
- Established computational pipeline for analysis of DNase-seq, ChIP-seq and mRNA-seq data

- Generated the first set of mRNA-seq data for AR activation by the ligand R1881 in LNCaP prostate cancer cells; this data will be made publicly available
- Uncovered novel dynamics of the chromatin template with regards to AR activation that stand in distinct contrast to the dynamics recently reported for other nuclear receptors.
- Specifically discovered that
    - Quantitative analysis of DNase-seq changes, as opposed to viewing chromatin as "open" or "closed", reveals unique insight into cofactors such as SP1 that may be involved in AR function
    - AR activation generally leads to increases in chromatin accessibility, especially in intergenic and intronic regions of the genome
    - AR binding does not just target regions of chromatin accessible to DNaseI prior to AR activation (in contrast to the GR). Rather, AR binding is often associated with an increase in chromatin accessibility
    - These increases in chromatin accessibility are associated with AR-mediated transcriptional changes
    - Uncovered DNaseI footprinting evidence of a possible monomeric AR-DNA interaction, which until now has only been speculated upon based on ChIP-seq motif analysis
- Presented findings at 2 national meetings with 1 poster presentation and 1 podium talk
- Submitted manuscript of findings that is under revision

Reportable Outcomes

Manuscripts

A manuscript entitled "Chromatin Accessibility reveals insights Into Androgen Receptor Activation and Transcriptional Specificity" was submitted on May 1, 2012 to the journal *Genome Biology*. We received reviews back on June 26, 2012. These reviews were largely favorable, but require significant response. We are currently preparing a response to the reviewer issues.

Abstracts and Presentations

1. Chromatin accessibility reveals insight into androgen receptor activation and transcriptional specificity. Presented as poster at AACR Special Conference on Advances in Prostate Cancer Research in Orlando, FL in February 2012.
2. Chromatin accessibility reveals insight into androgen receptor activation and transcriptional specificity. Presented as minisymposium podium presentation at AACR Annual Meeting in Chicago, IL in April 2012.

Informatics

Multiple high-quality and deeply sequence data sets have been generated in this first year of work that will be of use for both our research and that of the scientific community.

DNase-seq: We have generated multiple technical and biological replicates of DNase-seq data for LNCaP cells before and after AR activation by the widely-used synthetic ligand R1881 for 12 hours. This represents the only such data set available, and is now publicly available.

mRNA-seq: We have also generated multiple technical and biological replicates of mRNA-seq data matching the same cellular growth conditions as the DNase-seq data. This provides a powerful data set to relate chromatin accessibility to AR-mediated transcriptional changes, and is the only such mRNA-seq data set that we are aware of. This data will be made publicly available upon publication of our manuscript.

Conclusion

The AR is a transcription factor and a primary driver of prostate cancer. Understanding the key determinants of its transcriptional specificity remains a critical issue. By integrating analysis of DNase-seq data with AR ChIP-seq and mRNA-seq, we showed that AR activation induced genome-wide changes in chromatin structure that were associated with AR binding and transcriptional response and uncovered multiple modes of AR utilization of its DNA recognition motif. Although a subset of AR binding occurs in qualitatively poised chromatin exhibiting nucleosome depletion prior to hormone treatment, we demonstrated that AR binding is consistently associated with a quantitatively significant increase in DNase-seq signal suggesting stabilization of chromatin remodeling.

In the first year of this award, we have developed the computational expertise to handle various epigenetic and genetic high-throughput data as well as integrate them in a meaningful way to understand AR biology in prostate cancer. This work established the necessary framework to perturb AR function and elucidate the effects of the perturbation on chromatin structure, transcription and cellular phenotype.

References

1. Jemal A, Siegel R, Xu J, *et al.* Cancer statistics, 2010. CA Cancer J Clin 2010; 60:277-300.
2. Scher HI, Halabi S, Tannock I, *et al.* Design and end points of clinical trials for patients with progressive prostate cancer and castrate levels of testosterone: recommendations of the Prostate Cancer Clinical Trials Working Group. J Clin Oncol 2008; 26:1148-1159.
3. Visakorpi T, Hyytinen E, Koivisto P, *et al.* In vivo amplification of the androgen receptor gene and progression of human prostate cancer. Nat Genet 1995; 9:401-406.
4. Taplin ME, Bubley GJ, Ko YJ, *et al.* Selection for androgen receptor mutations in prostate cancers treated with androgen antagonist. Cancer Res 1999; 59:2511-2515.
5. Chen CD, Welsbie DS, Tran C, *et al.* Molecular determinants of resistance to antiandrogen therapy. Nat Med 2004; 10:33-39.
6. Tomlins SA, Mehra R, Rhodes DR, *et al.* Integrative molecular concept modeling of prostate cancer progression. Nat Genet 2007; 39:41-51.
7. Wang Q, Li W, Zhang Y, *et al.* Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. Cell 2009; 138:245-256.
8. Stanbrough M, Bubley GJ, Ross K, *et al.* Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. Cancer Res 2006; 66:2815-2825.
9. Taylor BS, Schultz N, Hieronymus H, *et al.* Integrative genomic profiling of human prostate cancer. Cancer Cell 2010; 18:11-22.

10. Taplin ME, Bubley GJ, Shuster TD*, et al.* Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer. N Engl J Med 1995; 332:1393-1398.
11. Veldscholte J, Ris-Stalpers C, Kuiper GG*, et al.* A mutation in the ligand binding domain of the androgen receptor of human LNCaP cells affects steroid binding characteristics and response to anti-androgens. Biochem Biophys Res Commun 1990; 173:534-540.
12. He HH, Meyer CA, Chen MW*, et al.* Differential DNase I hypersensitivity reveals factor-dependent chromatin dynamics. Genome Res 2012;
13. Glass CK, Rosenfeld MG. The coregulator exchange in transcriptional functions of nuclear receptors. Genes Dev 2000; 14:121-141.
14. Gregory CW, He B, Johnson RT*, et al.* A mechanism for androgen receptor-mediated prostate cancer recurrence after androgen deprivation therapy. Cancer Res 2001; 61:4315-4319.
15. Craft N, Shostak Y, Carey M*, et al.* A mechanism for hormone-independent prostate cancer through modulation of androgen receptor signaling by the HER-2/neu tyrosine kinase. Nat Med 1999; 5:280-285.
16. Gioeli D, Ficarro SB, Kwiek JJ*, et al.* Androgen receptor phosphorylation. Regulation and identification of the phosphorylation sites. J Biol Chem 2002; 277:29304-29314.
17. Kumar-Sinha C, Tomlins SA, Chinnaiyan AM. Recurrent gene fusions in prostate cancer. Nat Rev Cancer 2008; 8:497-511.
18. Gopalan A, Leversha MA, Satagopan JM*, et al.* TMPRSS2-ERG gene fusion is not associated with outcome in patients treated by prostatectomy. Cancer Res 2009; 69:1400-1406.
19. Carver BS, Tran J, Gopalan A*, et al.* Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. Nat Genet 2009; 41:619-624.
20. Zong Y, Xin L, Goldstein AS*, et al.* ETS family transcription factors collaborate with alternative signaling pathways to induce carcinoma from adult murine prostate cells. Proc Natl Acad Sci U S A 2009; 106:12465-12470.
21. Iljin K, Wolf M, Edgren H*, et al.* TMPRSS2 fusions with oncogenic ETS factors in prostate cancer involve unbalanced genomic rearrangements and are associated with HDAC1 and epigenetic reprogramming. Cancer Res 2006; 66:10242-10246.
22. Bjorkman M, Iljin K, Halonen P*, et al.* Defining the molecular action of HDAC inhibitors and synergism with androgen deprivation in ERG-positive prostate cancer. Int J Cancer 2008; 123:2774-2781.
23. Chng KR, Chang CW, Tan SK*, et al.* A transcriptional repressor co-regulatory network governing androgen response in prostate cancers. EMBO J 2012;
24. Yu J, Mani RS, Cao Q*, et al.* An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. Cancer Cell 2010; 17:443-454.
25. Mani RS, Tomlins SA, Callahan K*, et al.* Induced chromosomal proximity and gene fusions in prostate cancer. Science 2009; 326:1230.
26. Lin C, Yang L, Tanasa B*, et al.* Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell 2009; 139:1069-1083.

27. Mosquera JM, Perner S, Genega EM*, et al.* Characterization of TMPRSS2-ERG fusion high-grade prostatic intraepithelial neoplasia and potential clinical implications. Clin Cancer Res 2008; 14:3380-3385.
28. Perner S, Mosquera JM, Demichelis F*, et al.* TMPRSS2-ERG fusion prostate cancer: an early molecular event associated with invasion. Am J Surg Pathol 2007; 31:882-888.
29. Elgin SC. The formation and function of DNase I hypersensitive sites in the process of gene activation. J Biol Chem 1988; 263:19259-19262.
30. Boyle AP, Davis S, Shulha HP*, et al.* High-resolution mapping and characterization of open chromatin across the genome. Cell 2008; 132:311-322.
31. Crawford GE, Davis S, Scacheri PC*, et al.* DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. Nat Methods 2006; 3:503-509.
32. Hesselberth JR, Chen X, Zhang Z*, et al.* Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat Methods 2009; 6:283-289.
33. Biddie SC, John S, Sabo PJ*, et al.* Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. Mol Cell 2011; 43:145-155.
34. John S, Sabo PJ, Thurman RE*, et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nat Genet 2011; 43:264-268.
35. Boyle AP, Guinney J, Crawford GE*, et al.* F-Seq: a feature density estimator for high-throughput sequence tags. Bioinformatics 2008; 24:2537-2538.
36. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010; 26:139-140.
37. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25:1754-1760.
38. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 2009; 25:1105-1111.
39. Massie CE, Lynch A, Ramos-Montoya A*, et al.* The androgen receptor fuels prostate cancer by regulating central metabolism and biosynthesis. EMBO J 2011; 30:2719-2733.
40. Andreu-Vieyra C, Lai J, Berman BP*, et al.* Dynamic nucleosome-depleted regions at androgen receptor enhancers in the absence of ligand in prostate cancer cells. Mol Cell Biol 2011; 31:4648-4662.
41. Berman BP, Frenkel B, Coetzee GA*, et al.* Androgen receptor responsive enhancers are flanked by consistently-positioned H3-acetylated nucleosomes. Cell Cycle 2010; 9:2249-2250.
42. Song L, Zhang Z, Grasfeder LL*, et al.* Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res 2011; 21:1757-1767.
43. Safe S, Abdelrahim M. Sp transcription factor family and its role in cancer. Eur J Cancer 2005; 41:2438-2448.
44. Boyle AP, Song L, Lee BK*, et al.* High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. Genome Res 2011; 21:456-464.
45. Pique-Regi R, Degner JF, Pai AA*, et al.* Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res 2011; 21:447-455.

46. Shaffer PL, Jivan A, Dollins DE*, et al.* Structural basis of androgen receptor binding to selective androgen response elements. Proc Natl Acad Sci U S A 2004; 101:4758-4763.
47. Tang Q, Chen Y, Meyer C*, et al.* A comprehensive view of nuclear receptor cancer cistromes. Cancer Res 2011; 71:6940-6947.
48. Wang Q, Li W, Liu XS*, et al.* A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. Mol Cell 2007; 27:380-392.
49. Berger R, Febbo PG, Majumder PK*, et al.* Androgen-induced differentiation and tumorigenicity of human prostate epithelial cells. Cancer Res 2004; 64:8867-8875.

Appendices

None