Vitamin D receptor cistrome-transcriptome analyses establishes quantitatively distinct receptor genomic interactions in African American prostate cancer regulated by BAZ1A.

Manjunath Siddappa^{1ab}, Shahid Hussain^{1a}, Sajad A. Wani¹, Hancong Tang¹, Jaimie S. Gray¹, Hedieh Jafari¹, Hsueh Chang Wu¹, Mark D. Long², Isra Elhussain³, Balasubramanyam Karanam³, Honghe Wang³, Rebecca Morgan⁴, Gary Hardiman^{4,5}, Isaacson B. Adelani⁶, Solomon O. Rotimi⁶, Adam R Murphy⁷, Larisa Nonn⁸, Melissa B Davis⁹, Rick A Kittles¹⁰, Chanita Hughes Halbert^{11,12}, Lara E. Sucheston-Campbell^{13,14}, Clayton Yates³, Moray J. Campbell¹.

¹Pharmaceutics and Pharmaceutical Chemistry, College of Pharmacy, The Ohio State University, Columbus, OH 43210

²Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263

³Department of Biology and Center for Cancer Research, Tuskegee University, Tuskegee, AL 36088,

⁴School of Biological Sciences, Institute for Global Food Security, Queen's University Belfast, Belfast, United Kingdom.

⁵Department of Medicine, Medical University of South Carolina, Charleston, SC 29425

⁶Department of Biochemistry, Covenant University, KM. 10 Idiroko Road, Canaan Land, 112104, Ota, Ogun State, Nigeria

⁷Department of Urology, Northwestern Medicine, Chicago, Illinois.

⁸Department of Pathology, University of Illinois at Chicago, Chicago, Illinois, USA.

⁹Department of Surgery, Weill Cornell Medicine, 420 E 70th Street, New York City, NY, 10021, USA. mbd4001@med.cornell.edu.

¹⁰Division of Health Equities, Department of Population Sciences, City of Hope, Duarte, 91010, CA, USA

¹¹Department of Population and Public Health Sciences, University of Southern California, Los Angeles, CA 30089

¹²Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 20089

¹³Division of Pharmacy Practice and Science, College of Pharmacy, The Ohio State University, Columbus, OH 43210.

¹⁴Department of Veterinary Biosciences, College of Veterinary Medicine, The Ohio State University, Columbus, OH 43210

^aThese authors contributed equally to the study

^bCurrent address: Dept. of Veterinary Services and Animal Husbandry, Karnataka, India, 580001

Author emails:

Manjunath Siddappa; smanju712@gmail.com
 Shahid Hussain Sofi; Sofi.2@osu.edu
 Sajad A. Wani; wani.16@osu.edu
 Hancong Tang; tang.889@buckeyemail.osu.edu
 Hedieh Jafari; jafari.15@buckeyemail.osu.edu
 Jaimie S. Gray; gray.1291@buckeyemail.osu.edu
 Hsueh Chang Wu; wu.3606@buckeyemail.osu.edu
 Mark D. Long; Mark.Long@RoswellPark.org
 Isra Elhussin; ielhussin0014@tuskegee.edu
 Balasubramanyam Karanam; bkaranam@tuskegee.edu
 Honghe Wang; hwang@tuskegee.edu
 Rebecca Morgan; rmorgan21@qub.ac.uk
 Gary Hardiman; g.hardiman@qub.ac.uk
 Isaacson B. Adelani; isaacson909@gmail.com
 Solomon O. Rotimi; ola.rotimi@covenantuniversity.edu.ng
 Adam R Murphy; a-murphy2@northwestern.edu
 Larisa Nonn; lnonn@uic.edu
 Melissa B Davis; mbd4001@med.cornell.edu
 Rick A Kittles; rkittles@coh.org
 Chanita Hughes-Halbert; hughesha@usc.edu
 Lara E. Sucheston-Campbell; lara.sucheston-campbell@roche.com
 Clayton Yates; cyates@tuskegee.edu
 Moray J. Campbell; campbell.1933@osu.edu

ABSTRACT

Background. African American (AA) prostate cancer (PCa) appears uniquely sensitive to $1\alpha,25(\text{OH})_2\text{D}_3$ signaling, compared to European American (EA) PCa, but the extent and impact of vitamin D receptor genomic functions remain poorly defined.

Results. A panel of EA and AA prostate epithelial cells (EA: HPr1-AR, LNCaP, AA: RC43N, RC43T, RC77N, RC77T) were analyzed with RIME to reveal the cell-specific composition of the VDR-complex. $1\alpha,25(\text{OH})_2\text{D}_3$ -dependent ATAC-Seq revealed the greatest impact on nucleosome positioning in RC43N and RC43T, with gain of nucleosome-free at enhancer regions. VDR ChIP-Seq identified stronger and more frequent VDR binding in RC43N and RC43T that was enriched for a larger and distinct motif repertoire, than EA cells. VDR binding significantly overlapped with core circadian rhythm transcription factors in AA cell line models. RNA-Seq also revealed significantly stronger $1\alpha,25(\text{OH})_2\text{D}_3$ dependent VDR transcriptional responses enriched for circadian rhythm and inflammation networks in AA cells. Whilst RC43N was most responsive, RC43T displayed distorted responses. Significantly reduced *BAZ1A/SMARCA5* in AA PCa samples was identified, and restored BAZ1A expression uniquely and significantly increased $1\alpha,25(\text{OH})_2\text{D}_3$ -regulated VDR targets in AA cells. These VDR-dependent cistrome-annotated genes were also uniquely and most significantly identified in three cohorts of AA PCa patients.

Conclusion. These data suggest VDR transcriptional control in the prostate is more potent and dynamic in AA men, and primed to govern inflammatory and circadian pathways. Reduced BAZ1A/SMARCA5 expression and/or reduced environmentally-regulated serum vitamin D₃ levels suppress these actions. Therefore, the VDR axis lies at the cross-roads of biopsychosocial processes including stress responses, access to quality early detection and treatment, social determinants and that collectively contribute to PCa health disparities.

Keywords: Vitamin D Receptor, Prostate Cancer, Health Disparities, cistrome, transcriptome, SWI/SNF (SWItch/Sucrose Non-Fermentable), BAZ1A, SMARCA5

Running title: VDR and African American prostate cancer

BACKGROUND

The challenges of accurate early diagnosis of aggressive prostate tumors are amplified by a significant and devastating racial health disparity facing African American (AA) prostate cancer (PCa) patients. PCa in AA men is more aggressive and occurs at a younger age than in European American (EA) counterparts[1]. Current understanding of the drivers of PCa health disparities leverages genetic [2-4] and epigenetic [5-10] factors, combined with a range of biopsychosocial processes. The combined interactions of these intrinsic and extrinsic factors appear to drive uniquely aggressive disease in AA men. That there are biological drivers of PCa health disparities is supported strongly by the increased incidence of *TMPRSS2* and *ETS* fusion genes, which can act as cancer-drivers[11, 12], in EA PCa patients compared to AA counterparts [8, 13, 14].

To identify biological drivers of AA PCa we have focused on the vitamin D receptor (VDR), as this nuclear hormone receptor is implicated in cancer health disparities[15-17]. Ultraviolet B (UVB) radiation in the skin converts 7-dehydrocholesterol to cholecalciferol which is metabolized further to active $1\alpha,25(\text{OH})_2\text{D}_3$. However, UVB radiation also degrades folic acid, which is also essential, and during evolution and adaptation a strong inverse correlation between skin pigmentation and latitude has arisen [18, 19]. In more recent human history relatively rapid dispersion and lifestyle changes have resulted in many individuals living in an UVB environment which differ from those of their ancestors.

Significant associations have been observed between either altered VDR signaling or low serum vitamin D₃ levels and cancer incidence and progression risks amongst AA PCa patients [15-17, 20-32]. Furthermore, vitamin D₃ supplementation has been evaluated in large-scale randomized trials to define individual health benefits[33, 34]. In the VITAL cohort of approximately 26,000 men and women vitamin D₃ supplementation had no overall impact on cancer incidence, although AA participants experienced a suggestive 23% ($p=0.07$) reduction in cancer risk, indicating that larger cohorts maybe more informative in terms of defining the benefits of supplementation for these men [35-37]. Furthermore links between VDR signaling and anti-inflammatory actions appear strongest in AA PCa patients compared to EA counterparts [26]. Specifically, vitamin D₃ supplementation prior to radical prostatectomy significantly altered tumor expression of genes associated with inflammation in the prostates of AA compared to EA patients[38].

Aside from calcemic functions, the VDR regulates gene networks that control cell cycle progression, promote cell differentiation, and immunomodulatory events (reviewed in[39]). More

recently, independently of genomic ancestry, VDR signaling in has been associated with regulation of the circadian rhythm[40], reflecting a wider function for nuclear receptors to regulate this key process (reviewed in[41]). Together these data suggest that AA men may have a unique vulnerability to inadequate VDR signaling, which was potentially overlooked in the clinical evaluation of vitamin D₃ analogs in PCa cohorts that largely consisted of EA men[42, 43].

Therefore, we sought to define VDR signaling capacity in EA and AA prostate cell models and clinical cohorts. We utilized EA and AA prostate cell models with confirmed ancestry; namely non-malignant EA prostate cells (HPr1AR)[44, 45], EA PCa (LNCaP), non-malignant prostate cells AA (RC43N, RC77N) and AA PCa (RC43T, RC77T)[46, 47]. Using these models, we defined the basal and 1 α ,25(OH)₂D₃-regulated VDR protein interactome (RIME), the VDR cistrome (ChIP-Seq and ATAC-Seq) and the VDR transcriptome (RNA-Seq). These interactome-cistrome-transcriptome relationships were associated with outcomes in several clinical cohorts; 1) in AA and EA men with high grade prostatic intraepithelial neoplasia (HGPIN) who were at risk of progression to PCa; 2); a PCa chemoprevention trial where AA and EA men with localized PCa were supplemented with vitamin D₃ and 3) a PCa cohort of AA and EA men with gene expression data, clinical data and measured serum vitamin D₃ levels. These approaches combine to reveal that AA cells displayed more potent, and qualitatively and quantitatively different VDR signaling than EA cells, and reduced expression of in AA PCa of members of the SWI/SNF ATP-dependent chromatin remodeling complex, including *BAZ1A* and *SMARCA5* suppressed VDR-dependent gene expression.

Together these studies support the concept that AA prostate cells display greater VDR genomic signaling capacity than EA counterparts and that disruption through altered expression of the *BAZ1A*/*SMARCA5* chromatin remodeling complex primes cells to lose the ability to control immunomodulatory and circadian signaling.

RESULTS

The composition of the VDR complex and response to $1\alpha,25(\text{OH})_2\text{D}_3$ differs significantly across prostate epithelial cells. We measured the anti-proliferative responses towards $1\alpha,25(\text{OH})_2\text{D}_3$ across the six EA and AA cell models. Using optimal seeding conditions all cells had approximately equal doubling times (data not shown). **Figure 1A** demonstrates that the cancer cell lines (LNCaP, RC43T and RC77T) had more potent anti-proliferative responses towards $1\alpha,25(\text{OH})_2\text{D}_3$ than the non-malignant counterparts. Only RC43T and LNCaP cells form colonies and RC43N and HPr1AR do not. LNCaP displayed a clear $1\alpha,25(\text{OH})_2\text{D}_3$ -mediated inhibition of clonal proliferation, RC43T colony number was only stimulated by $1\alpha,25(\text{OH})_2\text{D}_3$ treatment (**Figure 1B**).

Next, we measured how the VDR complex differed between the four cell models. VDR protein levels were generally elevated following $1\alpha,25(\text{OH})_2\text{D}_3$ treatment, which was most pronounced in RC43N cells (**Figure 1C**). Subsequently, we measured VDR interacting proteins in the presence of chromatin in the basal and $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated states (100 nM, 4h) using RIME [48]. The number of positively enriched proteins in the basal state were; HPr1AR, n = 116; LNCaP, n = 135; RC43N, n = 115; RC43T, n = 245. Following addition of $1\alpha,25(\text{OH})_2\text{D}_3$ in HPr1AR and LNCaP the number of positively enriched proteins remained broadly constant (n = 119 and n = 110, respectively). However, more proteins were enriched in RC43N (n = 198) and profoundly fewer were enriched in RC43T (n = 15) (**Supplementary Table 1, Supplementary Figure 1B**).

The enriched protein data were classified as to whether they were Coactivators (CoA), corepressors (CoR), mixed function coregulators (Mixed), Transcription factors (TF) or other proteins. **Supplementary Table 1** indicates the number of proteins by class, and the most significant class member, and suggests both commonality and diversity in VDR interacting partners across cell lines. For example, across all cells the most enriched TF in either the basal or $1\alpha,25(\text{OH})_2\text{D}_3$ -treated state was either VDR or its heterodimer partners RXR α and RXR β , whereas the most enriched cofactors or other TFs varied by cell line. Other commonly enriched classes of proteins included heat shock proteins (e.g. HSP90AA1), RNA binding proteins (e.g. RBM12B), DNA repair enzymes (e.g. XRCC family members) and DNA helicases (e.g. DDX3X).

To measure the extent of the differences in the VDR complex we calculated the significant differences between enrichments in the basal and $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated states across cells. This revealed significant differences of proteins enriched in either the basal and $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated

states comparing RC43N to HPr1AR, and RC43T to LNCaP (**Figure 1D**). For example, $1\alpha,25(\text{OH})_2\text{D}_3$ treatment of RC43N cells enriched HDAC2, TRIM29 and SMARCA4 in the VDR complex.

Cell	Rx	Direct	Unique.Cell	class	Number	MostSignificant
RC43N.HPr1AR	EtOH.delta	P	RC43N.B	CoR	1	TRIM29
RC43N.HPr1AR	EtOH.delta	P	RC43N.B	other	26	YWHAE
RC43N.HPr1AR	EtOH.delta	N	RC43N.B	CoA	3	DDX39B
RC43N.HPr1AR	EtOH.delta	N	RC43N.B	TF	3	ZNF22
RC43N.HPr1AR	EtOH.delta	N	RC43N.B	CoR	5	KHDRBS1
RC43N.HPr1AR	EtOH.delta	N	RC43N.B	other	42	HNRNPA1
RC43N.HPr1AR	D3.delta	P	RC43N.D3	TF	6	STAT1
RC43N.HPr1AR	D3.delta	P	RC43N.D3	other	76	IARS
RC43N.HPr1AR	D3.delta	P	RC43N.D3	CoR	2	XPO1
RC43N.HPr1AR	D3.delta	P	RC43N.D3	CoA	3	TOP2A
RC43N.HPr1AR	D3.delta	N	RC43N.D3	CoR	4	NOC2L
RC43N.HPr1AR	D3.delta	N	RC43N.D3	CoA	4	BAZ1B
RC43N.HPr1AR	D3.delta	P	RC43N.D3	Mixed	1	HMGB1
RC43N.HPr1AR	D3.delta	N	RC43N.D3	Mixed	2	NCOA5
RC43N.HPr1AR	D3.delta	N	RC43N.D3	TF	7	ZNF638
RC43N.HPr1AR	D3.delta	N	RC43N.D3	other	90	RPS11
RC43T.LNCaP	EtOH.delta	P	RC43T.B	other	46	TGFB1
RC43T.LNCaP	EtOH.delta	P	RC43T.B	TF	4	SAFB2
RC43T.LNCaP	EtOH.delta	P	RC43T.B	CoA	1	RBMXL1
RC43T.LNCaP	EtOH.delta	P	RC43T.B	CoR	1	XPO1
RC43T.LNCaP	EtOH.delta	P	RC43T.B	Mixed	1	NCOA5
RC43T.LNCaP	EtOH.delta	N	RC43T.B	CoA	3	DDX39B
RC43T.LNCaP	EtOH.delta	N	RC43T.B	other	6	NOP56
RC43T.LNCaP	EtOH.delta	N	RC43T.B	CoR	2	XRCC6
RC43T.LNCaP	D3.delta	P	RC43T.D3	other	2	TGFB1

Table 1: Significantly and uniquely differentially enriched VDR-associated proteins. RIME analyses captured VDR interacting proteins and significant differences were measured between cell types and in either the basal state of following $1\alpha,25(\text{OH})_2\text{D}_3$ treatment (100 nM, 4h). Significant ($p_{\text{adj}} < .1$) differentially and uniquely enriched proteins in each cell and treatment condition were classified either as a Coactivator (CoA), Corepressor (CoR), Mixed function coregulator (Mixed) or transcription factor (TF), and the most significant member of each class in each condition is indicated.

From the differentially enriched proteins, we next filtered those that are unique to each cell and treatment. **Table 1** summarizes the number of significant and unique gain and loss of protein enrichment by class, and the most significant class member, which reemphasizes the diversity in VDR interacting partners across cell lines. For example, between RC43N and HPr1AR there were a large number of uniquely enriched VDR-associated proteins including in the basal state, the CoR TRIM29 and following $1\alpha,25(\text{OH})_2\text{D}_3$ treatment the CoR NOC2L is reduced and various cofactors that promote

transcription associate including the TF STAT1 and DNA topoisomerase TOP2A. By contrast, there were fewer unique differences between RC43T and LNCaP cells and $1\alpha,25(\text{OH})_2\text{D}_3$ treatment only uniquely changed 2 proteins between RC43T and LNCaP cells including TGFB1. Finally, we filtered these to the most pronounced changes and grouped them by coactivator and corepressor complexes (**Supplementary Figure 2, left, right**). Also, proteins associated with circadian rhythm were examined. Notably, components of the circadian rhythm control were differentially enriched across cells, with gains of in RC43T compared to LNCaP and loss in RC43N compared to HPr1AR cells, and a similar reciprocal enrichment with NONO in RC43N and loss in RC43T comparisons (**Figure 1E**).

$1\alpha,25(\text{OH})_2\text{D}_3$ -dependent nucleosome positioning is distinct across cells. To understand the impact of the VDR complex on nucleosome positioning we undertook vehicle or $1\alpha,25(\text{OH})_2\text{D}_3$ (100 nM, 4h) treatments and isolated chromatin for ATAC-Seq. Analyses of how $1\alpha,25(\text{OH})_2\text{D}_3$ impacted the distribution of nucleosome-free (NF) mononucleosome (mono) (**Figure 2A, left, right**) and regions, revealed the greater $1\alpha,25(\text{OH})_2\text{D}_3$ -dependent impact in RC43N and RC43T than either HPr1AR or LNCaP. For example, in RC43N there were ~11000 significant changes in $1\alpha,25(\text{OH})_2\text{D}_3$ -induced NF regions; 99.8% of which were gain in NF regions compared to the basal state. Similarly, 98% of the significant NF changes in HPr1AR were increases, but there were only ~950 regions and a comparable response was also seen in LNCaP, although there were only ~300 NF regions. There were no significant $1\alpha,25(\text{OH})_2\text{D}_3$ -regulated changes in mono regions in HPr1AR and RC43N, but there was a gain of ~1800 mono regions in LNCaP cells. In RC43T, by contrast, $1\alpha,25(\text{OH})_2\text{D}_3$ -induced changes in NF and mono regions were predominantly a loss compared to controls. Thus, there were ~1900 NF and ~800 mono-nucleosome regions significantly impacted by $1\alpha,25(\text{OH})_2\text{D}_3$ and these were mostly regions with a negative fold change compared to vehicle controls. Furthermore, the $1\alpha,25(\text{OH})_2\text{D}_3$ -regulated NF regions were significantly overlapped between RC43N and both RC43T and HPr1AR ($p < 4.1\text{e-}230$), whereas the LNCaP regions were distinct. Given that the shared NF regions between RC43N and RC43T are induced by $1\alpha,25(\text{OH})_2\text{D}_3$ in RC43N but lost in RC43T, this further supports the concept that VDR function is distinct between these cells. Next, we analyzed how the NF and mono basal and $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated genomic partitions were distributed and enriched in prostate ChromHMM defined regions. The ChromHMM algorithm combines ChIP-Seq data to define epigenetic states and are publicly available for LNCaP cells[49]. The $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated NF and mono regions significantly enriched most frequently across Promoter regions (**Table 2**).

Cell	RX	nuc	ChromHMM	logPval	Threshold	Direction
LNCaP	D3	mono	Promoter	315.86	Significant	Gain

RC43N	D3	NF	Promoter	315.86	Significant	Gain
RC43T	D3	mono	Promoter	315.86	Significant	Loss
HPr1AR	D3	NF	Promoter	315.51	Significant	Gain
RC43T	D3	NF	Promoter	168.63	Significant	Loss
LNCaP	D3	NF	Promoter	46.48	Significant	Gain
RC43N	D3	NF	Transcribed	9.95	Significant	Gain
RC43T	D3	NF	Transcribed	8.24	Significant	Loss
RC43N	D3	NF	Bivalent_Promoter	1.40	Significant	Gain

Table 2: Enrichment of $1\alpha,25(\text{OH})_2\text{D}_3$ -induced nucleosome free (NF) and mononucleosome (mono) ChromHMM regions. ATAC-Seq was undertaken following $1\alpha,25(\text{OH})_2\text{D}_3$ treatment (100 nM, 4h) or vehicle control. Reads were processed by ATACseqQC to nucleosome fractions, and then differentially $1\alpha,25(\text{OH})_2\text{D}_3$ enriched NF and mono free regions ($p_{\text{adj}} < .1$) were identified by csaw and were then overlapped with ChromHMM regions (ChromHMM) identified in LNCaP using bedtools and enrichment tested with a hypergeometric test (lower.tail = FALSE).

We mined these NF and mono regions for motif enrichment to illuminate how changes in nucleosome positioning may impact transcription factor binding [50]. Consistent with the fact that the greatest number of $1\alpha,25(\text{OH})_2\text{D}_3$ -induced NF regions were identified RC43N cells, there were also more significant motif enrichments (by p value and percentage coverage) in this cell line. In **Figure 2B** and **Supplementary Figure 3A** the TF families are ranked by the significance of enrichment. For example, in RC43N, the most significantly enriched motifs included AP (e.g., AP-1), bZIP (e.g., ATF3), along with SP, FOX and ETS family members. This was broadly reflected by HPr1AR and RC43T cells, with RC43T cells showing a loss of potential TF interactions. LNCaP cells, by contrast, displayed fewer significant associations. Interestingly, one of the most enriched motifs in LNCaP was the AR/Half-site, which was also only enriched in RC43T regions. Enrichment in mono regions included SP (e.g., SP2), C2H2-ZNF (e.g. KLF5) and ETS (e.g. ELK4) family members ((**Figure 2B**; **Supplementary Figure 3A**).

We calculated delta values comparing the differences between cells in motif enrichment in the $1\alpha,25(\text{OH})_2\text{D}_3$ -induced NF regions (**Figure 2C**). Compared to HPr1AR cells, RC43N displayed the most striking gains in enrichment of multiple TF families including bHLH and FOX family members. The bHLH motifs included multiple TFs for circadian rhythm such as bMAL1 and CLOCK, which were either most clearly or exclusively enriched compared to other cells (**Figure 2D**). Similarly, FOX family members were highly enriched in RC43N, although it is notable that the FOXA1-AR motif enrichment is most prominent in LNCaP cells (**Supplementary Figure 3B**).

The VDR cistrome is distinct across cells. The VDR cistrome was measured in cells treated with $1\alpha,25(\text{OH})_2\text{D}_3$ (100 nM, 6h) and VDR peaks ($\text{FDR} < 0.1$) identified compared to IgG controls. Significant basal VDR binding was identified in LNCaP, RC43N and RC43T cells but not HPr1AR, and in all cells

following $1\alpha,25(\text{OH})_2\text{D}_3$ -treatment (**Supplementary Table 2**). The basal peaks were highly overlapping (sharing at least 1 bp) with 305 shared sites across all three cells (**Figure 3A**). $1\alpha,25(\text{OH})_2\text{D}_3$ treatment reduced the number of VDR binding sites in LNCaP and RC43T; there were 773 and 1560 basal VDR peaks in LNCaP and RC43T cells, which reduced to 322 and 30, respectively. Within each cell there was also a large overlap between basal and $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated sites (**Supplementary Figure 4A**) and the proximal/distal distribution of the VDR cistrome was broadly constant within a cell type regardless of $1\alpha,25(\text{OH})_2\text{D}_3$ treatment (**Supplementary Figure 4B**).

Similar to the ATAC-Seq data, we analyzed how the basal and $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated genomic partitions were distributed and enriched in prostate ChromHMM defined regions (**Table 3**). The basal VDR cistrome across cells was commonly enriched in Transcribed regions and Bivalent Promoters and remained so following $1\alpha,25(\text{OH})_2\text{D}_3$ treatment. Other enrichments were cell specific, for example, Polycomb regions were only enriched in RC43T and LNCaP. Compared to the ATAC-Seq enrichments, which were frequent at Promoter regions, only basal VDR ChIP-Seq in LNCaP was enriched at Promoter regions, and only modestly so. Interestingly, this approach also revealed the significant impact of $1\alpha,25(\text{OH})_2\text{D}_3$ treatment. For example, in LNCaP the $\log_{10}(\text{p.adj})$ of the basal VDR enrichment in Bivalent Promoters was 46, and this was reduced to 6 in the presence of $1\alpha,25(\text{OH})_2\text{D}_3$, and similarly was 45 in RC43T in the basal state, and non-significant following $1\alpha,25(\text{OH})_2\text{D}_3$ treatment. By contrast the score in RC43N was broadly equivalent in the basal and $1\alpha,25(\text{OH})_2\text{D}_3$ states.

Cell.ChIP	ChromHMM	logPV	Threshold
LNCaP.VDR	Bivalent_Promoter	46.16	Significant
RC43T.VDR	Transcribed	45.74	Significant
RC43T.VDR	Bivalent_Promoter	40.84	Significant
RC43T.VDR	Polycomb	35.42	Significant
LNCaP.VDR	Polycomb	30.69	Significant
LNCaP.VDR	Transcribed	22.56	Significant
RC43N.VDR.D3	Bivalent_Promoter	20.51	Significant
RC43N.VDR	Bivalent_Promoter	15.49	Significant
LNCaP.VDR.D3	Bivalent_Promoter	6.37	Significant
HPr1AR.VDR.D3	Bivalent_Promoter	4.63	Significant
HPr1AR.VDR.D3	Transcribed	4.48	Significant
LNCaP.VDR	Poised_Enhancer	4.41	Significant
RC43N.VDR.D3	Transcribed	3.98	Significant
LNCaP.VDR.D3	Transcribed	3.07	Significant
RC43N.VDR	Transcribed	2.09	Significant
LNCaP.VDR	Promoter	1.56	Significant
RC43T.VDR	Poised_Enhancer	1.34	Significant

Table 3: Enrichment of VDR ChIP-Seq and ChromHMM regions. VDR ChIP-Seq was undertaken following $1\alpha,25(\text{OH})_2\text{D}_3$ treatment (100 nM, 4h) or vehicle control. Differential VDR binding sites compared to IgG controls ($p.\text{adj} < .1$) were identified by `csaw` and were then overlapped with ChromHMM regions (ChromHMM) identified in LNCaP using `bedtools` and enrichment tested with a hypergeometric test (`lower.tail = FALSE`).

Motif analyses categorized significantly enrichment of TF families across cells and treatments (**Table 4**). For example, the predicted VDR motif and the most significant nuclear receptor motif in LNCaP was also identified in RC43T. By contrast, the $\text{RAR}\alpha$ motif was the most significant nuclear receptor motif in HPr1AR and RC43N (**Figure 3B**). By contrast to the ATAC-Seq data, the enriched bHLH motifs did not include CLOCK, BMAL1, MITF and NFEs, but did include CEBP in all conditions. To gain further insight into how these motif enrichments changed by $1\alpha,25(\text{OH})_2\text{D}_3$ treatment and across cell types, we calculated the delta $\log(p.\text{adj})$ between groups, and visualized the most altered motifs (**Figure 3C**). This revealed, for example, in LNCaP, RC43N and RC43T cells that $1\alpha,25(\text{OH})_2\text{D}_3$ treatment reduced enrichment for various motifs including STAT5. In fact, only in RC43N cells did $1\alpha,25(\text{OH})_2\text{D}_3$ treatment increase motif enrichment notably for TEAD1 and TEAD3. Similarly, the circadian rhythm associated TF, NFAT, was more enriched in $1\alpha,25(\text{OH})_2\text{D}_3$ treated RC43N than HPr1AR.

Cell.ChIP	homer	MostSignificant	Log.Pval	class	Frequency
HPr1AR_D3	pred	RARA	403.9	NucRec	3
HPr1AR_D3	pred	STAT5	386.6	STAT	2
HPr1AR_D3	pred	BAPX1	241.7	Homeobox	8
LNCaP_EtOH	pred	VDR	781.6	NucRec	2
LNCaP_EtOH	pred	NKX2-2	760	Homeobox	3
LNCaP_EtOH	pred	KLF4	643.9	C2H2-ZNF	3
LNCaP_EtOH	pred	SMAD4	489.9	SMAD	2
LNCaP_EtOH	pred	STAT6	319.5	STAT	2
LNCaP_EtOH	known	ZNF189	142.9	Developmental	14
LNCaP_EtOH	known	STAT5	113.7	STAT	7
LNCaP_D3	pred	NKX2-3	311.6	Homeobox	4
LNCaP_D3	pred	STAT5	301.5	STAT	2
LNCaP_D3	pred	GATA5	162.7	GATA	2
RC43N_EtOH	pred	IRF4	1055	Interferon-reg	3
RC43N_EtOH	pred	NKX2-3	626.6	Homeobox	5
RC43N_EtOH	pred	SMAD4	533.6	SMAD	3
RC43N_EtOH	pred	KLF4	497.3	C2H2-ZNF	2
RC43N_EtOH	pred	SOX10	455	SRY-box	2
RC43N_EtOH	pred	FOXH1	324.2	FOX	3
RC43N_EtOH	known	ZNF189	196.2	Developmental	14
RC43N_EtOH	known	STAT5	145.8	STAT	7

RC43N_D3	pred	RARA	699.1	NucRec	2
RC43N_D3	pred	IRF4	639.2	Interferon-reg	3
RC43N_D3	pred	HDX	556.7	Homeobox	3
RC43N_D3	pred	ZNF528	317.4	ZNF	3
RC43N_D3	known	ZNF189	204	Developmental	14
RC43N_D3	known	STAT6	141.4	STAT	7
RC43N_D3	known	TEAD1	116.8	TEAD	5
RC43T_EtOH	pred	ZNF528	992.7	ZNF	2
RC43T_EtOH	pred	MAFK	760.6	bHLH	2
RC43T_EtOH	pred	NKX2-2	635.6	Homeobox	4
RC43T_EtOH	pred	FOXA2	474.5	FOX	2
RC43T_EtOH	known	ZNF189	162.2	Developmental	14
RC43T_EtOH	known	STAT5	135.1	STAT	7
RC43T_EtOH	pred	ZFP691	133.7	C2H2-ZNF	2

Table 4: Summary of motif enrichment by transcription family in the VDR dependent cistromes. Significant VDR binding sites were analyzed by Homer to identify enrichment of motifs, by transcription factor family enrichment. Within each cistrome the known and predicted (pred) motifs are arranged by $-\log(p_{adj})$ enrichment value for the most significant family member.

Next, VDR cistromes were analyzed for significant overlaps with TF and histone modifications contained in the CistromeDB collection (> 10,000 total ChIP-seq datasets, across > 1100 factors) using GIGGLE[51], which searches publicly available datasets within the Cistrome database and returns samples which contains peak sites co-incident with user-supplied data (**Supplementary Figure 5**); from these data we calculated the comboScore_{adj} as the negative $\log_{10}$ of the fisher.pval multiplied by the $\log_2$ of the odds.ratio. This approach reflected the motif analyses and provided additional strong evidence for VDR binding being co-incident with multiple TFs and histone modifications, and that the $1\alpha,25(\text{OH})_2\text{D}_3$ treatment dynamically altered many of these relationships. These analyses also revealed clear changes in numerous ZNF transcription factors (magenta) and a clear under-enrichment of bHLH members in the basal RC43N cells (**Supplementary Figure 5A**). There were significant overlaps with nuclear receptor ChIP-Seq data, including for PPARs, RARs, and VDR, which were all most significant in RC43N (**Supplementary Figure 5B**). For example, a combined score for overlapping with VDR ChIP-Seq (calculated as the odds ratio multiplied by the $-\log_{10}(\text{Fisher.p}_{adj})$), was 1111 and 1285 for RC43N in the basal and $1\alpha,25(\text{OH})_2\text{D}_3$ -treated cells in non-malignant prostate epithelial cells. These were much reduced in HPr1AR and LNCaP (4 and 71, respectively). Similarly, there were $1\alpha,25(\text{OH})_2\text{D}_3$ -dependent increases in enrichment for other nuclear receptors including AR, ER α , and RAR γ . Analyzing a canonical list of 32 transcriptional regulators of circadian rhythm revealed they significantly overlapped with VDR cistromes most significantly in RC43N. The most frequent enrichments were in both the basal and $1\alpha,25(\text{OH})_2\text{D}_3$ -treated RC43N cells, and indeed were amongst

the most significantly overlapping factors within any VDR cistrome, including NONO (comboScore.adj = 1745 RC43N 1 α ,25(OH) $_2$ D $_3$ -treated) (**Supplementary Figure 5C**).

Given that the median distance for enhancer to TSS is greater than 100 kb[52-54], the VDR cistromes (ATAC-Seq and ChIP-Seq) were related to genes within a 100 kb window up and downstream of genes. From these genes we measured which encoded for proteins that were also identified by RIME to interact with the VDR in either the basal or 1 α ,25(OH) $_2$ D $_3$ -stimulated state (**Table 5**). This suggested that ATAC-Seq data sets commonly annotated to genes that in many cases encoded for proteins that were positively enriched with the VDR. In HPr1AR, LNCaP and RC43N approximately equal numbers of basal and 1 α ,25(OH) $_2$ D $_3$ -stimulated proteins were enriched with VDR that were also targets of the cistrome data. However, this not the case in RC43T, which had a large number of overlaps between ATAC-Seq NF regions and the VDR interactome in the basal state but far fewer in the 1 α ,25(OH) $_2$ D $_3$ -stimulated state. Given the profound reduction of 1 α ,25(OH) $_2$ D $_3$ -stimulated NF regions in RC43T this may suggest a mechanism to limit the function of the VDR complex in this cell.

Cell.cistrome	RIME.Rx	class	Number	MostSignificant
HPr1AR.ATAC	D3	other	29	ITGB4
HPr1AR.ATAC	EtOH	other	24	YWHAQ
HPr1AR.ATAC	EtOH	TF	4	ARNTL2
HPr1AR.ATAC	EtOH	CoA	2	RBMXL1
HPr1AR.ATAC	D3	TF	2	SSRP1
HPr1AR.ATAC	EtOH	Mixed	1	CTBP2
HPr1AR.ATAC	EtOH	CoR	1	HSPA8
HPr1AR.ATAC	D3	CoR	1	HSPA8
HPr1AR.ATAC	D3	CoA	1	SAFB
LNCaP.ATAC	EtOH	other	3	PGD
LNCaP.ATAC	D3	other	3	PGD
LNCaP.ATAC	EtOH	TF	1	ENO1
LNCaP.ATAC	EtOH	CoR	1	NOC2L
LNCaP.ATAC	D3	TF	1	ENO1
LNCaP.ATAC	D3	CoR	1	NOC2L
RC43N.ATAC	D3	other	106	CLTC
RC43N.ATAC	EtOH	other	60	CLTC
RC43N.ATAC	D3	TF	10	VDR
RC43N.ATAC	D3	CoA	8	PTBP1
RC43N.ATAC	D3	CoR	7	XRCC6
RC43N.ATAC	EtOH	CoR	4	HSPA8
RC43N.ATAC	EtOH	TF	3	VDR
RC43N.ATAC	EtOH	CoA	2	PTBP1
RC43N.ATAC	D3	Mixed	1	HMGB1
RC43T.ATAC	EtOH	other	158	RBM12B

RC43T.ATAC	EtOH	TF	17	VDR
RC43T.ATAC	EtOH	CoA	15	SAFB
RC43T.ATAC	EtOH	CoR	8	XRCC5
RC43T.ATAC	D3	other	8	TGFBI
RC43T.ATAC	EtOH	Mixed	2	NCOA5
RC43T.ATAC	D3	TF	1	VDR
RC43T.ATAC	D3	CoR	1	XRCC6
LNCaP.ChIP	EtOH	other	1	TUBB4B
LNCaP.ChIP	EtOH	TF	1	GTF3C1
LNCaP.ChIP	D3	other	1	TUBB4B
RC43T.ChIP	EtOH	other	7	TNC
RC43T.ChIP	EtOH	CoR	1	TRIM29

Table 5: Genes annotated to $1\alpha,25(\text{OH})_2\text{D}_3$ -regulated nucleosome free regions and VDR binding sites overlap with VDR-interacting proteins. Genes were annotated to ATAC-Seq or ChIP-Seq regions within 100kB and those genes overlapped with the positively enriched proteins within the same cell background. Enriched genes were classified either as a Coactivator (CoA), Corepressor (CoR), Mixed function coregulator (Mixed), transcription factor (TF) or other, and the most significant member of each class in each condition is indicated.

The VDR transcriptome is larger in RC43N and RC43T than HPr1AR and LNCaP cells.

$1\alpha,25(\text{OH})_2\text{D}_3$ -dependent transcriptomic analyses were undertaken in HPr1AR, LNCaP, RC43N and RC43T using both a standard and small RNA library preparation for RNA-Seq (100 nM, 8h). Similarity and principal component analyses revealed that experimental conditions explained the majority of variation in expression (**Supplementary Figure 6**). Reflecting the size of the VDR cistromes, the number of $1\alpha,25(\text{OH})_2\text{D}_3$ regulated genes was larger in the AA than EA models. Differentially expressed genes in response to $1\alpha,25(\text{OH})_2\text{D}_3$ (**Figure 4A**; DEGs; FDR < 0.1, FC > 1.3) were determined in HPr1AR (teal), LNCaP (orange), RC43N (indigo) and RC43T (pink). The top most regulated genes included the well-established VDR target gene CYP24A1; CYP24A1 was not in the top 5 labelled genes in RC43N, although was in top 20 DEGs. Genes that had a $1\alpha,25(\text{OH})_2\text{D}_3$ -dependent NF region (red symbol) and/or a VDR binding peak (symbol with dark border) in the same cell background were more apparent in AA cells.

This was also noticeable when considering the DEGs across cells of different genomic ancestry, namely the difference in expression between AA vs. EA tumors exposed to $1\alpha,25(\text{OH})_2\text{D}_3$ (RC43T treated with $1\alpha,25(\text{OH})_2\text{D}_3$ compared to LNCaP treated with $1\alpha,25(\text{OH})_2\text{D}_3$) and difference in expression between AA vs. EA non-malignant cells exposed to $1\alpha,25(\text{OH})_2\text{D}_3$ (RC43N treated with $1\alpha,25(\text{OH})_2\text{D}_3$ compared to HPr1AR treated with $1\alpha,25(\text{OH})_2\text{D}_3$). Small RNA expression patterns were comparable across cells in terms of magnitude and frequency with ~20 short non-coding RNA, being regulated in each model, in a largely distinct manner (**Figure 4B**; DEGs; FDR < 0.1, FC > 1.3).

1 α ,25(OH) $_2$ D $_3$ -regulated miRNA that were also annotated to 1 α ,25(OH) $_2$ D $_3$ -dependent NF regions were identified in HPr1AR and RC43T and in the latter case included miR-99a.

We also used transcript-aware analyses to identify how 1 α ,25(OH) $_2$ D $_3$ -treatment induced differentially expressed transcripts (DETs). In each cell, 1 α ,25(OH) $_2$ D $_3$ -treatment induced significant DETs; HPr1AR, n=96; LNCaP, n=124; RC43N, n=136; RC43T, n=77, whereas the comparisons across cells were much larger (data not shown). Again, we categorized DETs functionally and in relationship to the cistrome data (**Table 6**). RC43N and RC43T cells displayed the highest proportion of DETs that were annotated to cistrome NF ATAC-Seq annotated genes. (RC43N, 84 out of 136; RC43T, 71 out of 77). Of these, 10 were TFs, and the most significant was ZNF83. Similarly, the coactivator SMARCA4 was an NF ATAC-Seq annotated gene and was a significant DET.

Cell	Rx	coreg	Cistrome	NumberDiffTranscript	MostSignificant
HPr1AR	D3	other	HPr1AR.VDR.ATAC	22	RPL22
HPr1AR	D3	Corepressor	HPr1AR.VDR.ATAC	1	BBS2
HPr1AR	D3	TF	HPr1AR.VDR.ATAC	1	ZNF18
LNCaP	D3	other	LNCaP.VDR.ATAC	2	SNHG12
LNCaP	D3	other	LNCaP.VDR.ChIP	2	LINC00623
RC43N	D3	other	RC43N.VDR.ATAC	66	SRP14
RC43N	D3	Coactivator	RC43N.VDR.ATAC	4	SMARCA4
RC43N	D3	TF	RC43N.VDR.ATAC	10	ZNF83
RC43N	D3	Mixed	RC43N.VDR.ATAC	2	ZMYND8
RC43N	D3	Corepressor	RC43N.VDR.ATAC	2	ZBTB21
RC43T	D3	Coactivator	RC43T.VDR.ATAC	7	STK16
RC43T	D3	other	RC43T.VDR.ATAC	50	PRKAA1
RC43T	D3	TF	RC43T.VDR.ATAC	5	ELL2
RC43T	D3	TF	RC43T.VDR.ChIP	1	ELL2
RC43T	D3	other	RC43T.VDR.ChIP	4	G3BP1
RC43T	D3	Coactivator	RC43T.VDR.ChIP	1	IFT74
RC43T	D3	Corepressor	RC43T.VDR.ATAC	3	BBS2

Table 6: Genes annotated to 1 α ,25(OH) $_2$ D $_3$ -regulated nucleosome free regions and VDR binding sites overlap with differentially expressed transcripts. Genes were annotated to ATAC-Seq or ChIP-Seq regions within 100kB and those genes overlapped with the differentially expressed transcripts, identified by salmon/DRIM-Seq. Overlapped genes were classified either as a Coactivator (CoA), Corepressor (CoR), Mixed function coregulator (Mixed), transcription factor (TF) or other, and the most significant member of each class in each condition is indicated.

Pre-ranked gene set enrichment analysis (GSEA) was undertaken using the Hallmarks, Chemical Perturbations and GO biological terms and underscored how divergent the 1 α ,25(OH) $_2$ D $_3$ -dependent gene expression patterns were between cell types (**Supplementary Figure 8**). HPr1AR and LNCaP cells displayed few significant enrichments (FDR < 0.1), and the only enriched term in LNCaP cells was enrichment of the vitamin D receptor pathway. By contrast, RC43N cells displayed

many positively enriched terms including Early Estrogen Response and negatively enriched terms including those associated with cell cycle control. Other modestly positively and negatively enriched gene sets were associated with inflammatory-response, AR and ER α signaling (positive), MYC signaling (negative), and several Circadian rhythm terms. RC43T displayed some similarities (e.g., Estrogen Response signaling and vitamin D pathway), but frequently the terms were either less significant, or enrichment was switched to a negative enrichment.

To test the possibility that indeed GSEA terms (e.g., **Supplementary Figure 9A**) were oppositely enriched, we calculated the delta between NES terms from RC43N and RC43T. Of the 211 terms that were significantly enriched in both RC43N and RC43T, 135 (64.0%) were oppositely regulated with a delta value of greater than 3 (**Figure 4C**). For example, Zhou Cell cycle genes enrichment score (NES) was -2.99 in RC43N, which switched to 1.28 in RC43T. Similarly, terms associated with estrogen signaling, cancer progression, MYC and SMARCE1-dependent targets were oppositely regulated. These findings were complemented by epigenetic landscape in silico deletion analysis, which measures enrichment of transcription factor targeting in the respective transcriptomes[55]. **Supplementary Figure 9B** displays a Euler diagram of the overlap of these transcription factors and demonstrates that although RC43N and RC43T contained distinct set members, but were also essentially overlapping with HPr1AR and LNCaP, supporting the concept that the response in RC43N and RC43T cells is most diverse.

Together these findings reveal that 1 α ,25(OH) $_2$ D $_3$ transcriptional signaling is more impactful in AA compared to EA prostate models, and that within AA cells, there is frequent gain and loss of enrichment of many pathways suggesting that the capacity of 1 α ,25(OH) $_2$ D $_3$ signaling is altered.

Integration of VDR dependent cistromes and transcriptomes reveals stronger gene regulation responses in RC43N and RC43T cells. In the first instance we examined how the cistrome data overlapped by comparing the ATAC-Seq and ChIP-Seq data for each cell and treatment combination. We measured the intersection of the basal and 1 α ,25(OH) $_2$ D $_3$ -stimulated VDR ChIP-Seq with the 1 α ,25(OH) $_2$ D $_3$ -stimulated mono and NF ATAC-Seq data. Remarkably, none of the cells displayed any significant overlap (a minimum of 1 bp) between ChIP- and 1 α ,25(OH) $_2$ D $_3$ -stimulated ATAC-Seq data. That is, although within each cell and condition there were individual genes annotated to have both a significant VDR binding site and a significant 1 α ,25(OH) $_2$ D $_3$ -stimulated NF region, these sites did not overlap. By contrast, in RC43N the basal and 1 α ,25(OH) $_2$ D $_3$ -stimulated VDR ChIP-Seq data

significantly overlapped with the separate basal and $1\alpha,25(\text{OH})_2\text{D}_3$ treated NF ATAC-Seq (i.e., not the differentially enriched $1\alpha,25(\text{OH})_2\text{D}_3$ -dependent NF regions) and the basal conditions only in LNCaP (**Supplementary Table 3**) supporting a role for VDR to bind to NF regions in these cells.

Next, each cistrome set was integrated with the transcriptome to identify relationships between either VDR binding or $1\alpha,25(\text{OH})_2\text{D}_3$ -induced NF regions, and gene expression, namely peak:gene relationships. To refine these analyses further, the VDR and cistromes were annotated to ChromHMM-defined chromatin states, and annotated to within a 100 kb of the nearest gene, including those that were members of the VDR-interactome as defined by BioGRID [56] (**Supplementary Table 4, 5**). Given there are more $1\alpha,25(\text{OH})_2\text{D}_3$ -dependent NF regions, the ATAC-Seq data was annotated to more genes. Within these genes it is interesting to note that RC43N NF regions in ChromHMM defined promoters annotate to the highest number of genes, followed by RC43T. It is also clear that many VDR.biogrid genes are associated with gain NF regions including *TP53* and *NCOR1* in RC43N. Similarly, in RC43T, the loss of NF regions annotates to many VDR BioGRID genes including *LCOR* and *RB*.

By contrast, the basal and $1\alpha,25(\text{OH})_2\text{D}_3$ -dependent VDR ChIP-Seq data were annotated to fewer genes, and largest number is in RC43N, where they are evenly distributed evenly between the different ChromHMM regions. Again, reflecting how RC43T cells may lose VDR-genome interactions in the presence of $1\alpha,25(\text{OH})_2\text{D}_3$, almost all peak:gene relationships are in the basal state and most are not annotated to ChromHMM regions. Finally, across the ChIP-Seq data, very few VDR.biogrid genes are identified with notably including *LCOR*.

We further reasoned that the VDR cistromes represent the sum potential for $1\alpha,25(\text{OH})_2\text{D}_3$ -dependent gene expression and therefore restricted VDR-dependent gene:peak relationships to genes detected by regular and small RNA-Seq expressed in the comparisons across AA to EA models (e.g. **Supplementary Figure 7**; AA.N.D3 (RC43N/ $1\alpha,25(\text{OH})_2\text{D}_3$ compared to HPr1AR/ $1\alpha,25(\text{OH})_2\text{D}_3$); AA.T.D3 (RC43T/ $1\alpha,25(\text{OH})_2\text{D}_3$ compared LNCaP/ $1\alpha,25(\text{OH})_2\text{D}_3$)). Restricting to these genes we then categorized peak:gene relationships for the ChromHMM-classified VDR cistromes and ChromHMM-classified ATAC NF cistromes (**Supplementary Figure 10A**). Next, we further categorized these peak:gene relationships by calculating frequency in 10kb bins up and downstream from DEGs (ChIP-Seq data; **Supplementary Figure 10B**, ATAC-Seq data; data not shown). RC43N gene:peak relationships were most frequent of the models being detected in all classifications in both the basal and $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated states. By contrast, in RC43T the gene:peak relationships were only

detected in the basal state. That is, the 30 $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated VDR binding sites in RC43T (**Supplementary Table 2**), annotated to 19 genes (**Supplementary Table 4**) but none of these were DEGs identified between RC43T/ $1\alpha,25(\text{OH})_2\text{D}_3$ compared to LNCaP/ $1\alpha,25(\text{OH})_2\text{D}_3$.

To test the significance of the VDR-dependent cistrome-transcriptome relationships we applied these classifications and built on the BETA method[57]. Specifically, the sum was calculated of VDR peak or NF scores within 10 kb bins around DEGs multiplied by the absolute gene-specific fold change for each DEG, and weighted to more proximal relationships; this approach was used to test how the summed VDR binding in defined genomic features significantly related to changes in gene expression.

In the first instance, for the ChIP-Seq data (categorized as EtOH-dependent, D3-dependent or shared EtOH.D3-dependent) were integrated with RNA-Seq, regardless of ChromHMM association. This revealed that the few genes in HPR1AR that were bound by $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated VDR were only identified in the positively regulated genes. All other comparisons were compared to RC43N cells, and revealed that in the basal peaks the scores for positively regulated genes in RC43T were significantly greater. This may suggest that basal VDR-bound genes in RC43N are less transcriptionally responsive than the same condition in RC43T. These analyses underscored that lack of a $1\alpha,25(\text{OH})_2\text{D}_3$ VDR-mediated response in RC43T (**Figure 5A**). This was also true considering only genes bound in the Transcribed regions (data not shown).

With more regions in the ATAC-Seq data, and only initiated by $1\alpha,25(\text{OH})_2\text{D}_3$, these relationships were considered in ChromHMM regions. VDR binding and impact across the cells was compared to RC43T and in several cases RC43N summed peak scores were significantly greater, for example Promoters and Poised and Active enhancers, and for negatively regulated genes in Transcribed regions (**Figure 5B**). Together these data support the concept that that VDR binding and $1\alpha,25(\text{OH})_2\text{D}_3$ associated NF regions in RC43N were most consistently associated with stronger patterns of gene expression than the other three cells.

The BAZ1A/SMARCA5 complex regulates $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated VDR responses. The cistrome-transcriptome studies support the concept that non-malignant RC43N prostate cells are significantly more sensitive to VDR-mediated gene regulation than either RC43T or the EA cells. The VDR binding in RC43N is co-incident with motifs for other nuclear receptors and other factors and including transcription factors that control circadian rhythm. By contrast to RC43N, the isogenic malignant counterpart RC43T, displays suppressed gene expression patterns that include disruption of

numerous transcriptional programs. Therefore, we exploited clinical cohorts to determine whether there was evidence for suppressed VDR signaling in AA PCa progression.

In the first instance, we screened a large panel of coregulators[58] in the DEGs between *TPR52* fusion positive and negative PCa in EA and AA patients in the TCGA-PRAD cohort. This identified 27 altered coregulators in AA patients, and from these, five were uniquely or more significantly altered in AA compared to EA *TPR52* fusion negative PCa (**Figure 6A**). The most altered coregulators included several known to interact with VDR signaling including *BAZ1A* (Bromodomain Adjacent to Zinc Finger Domain 1A), and *SMARCA5/WSTF* (SWI/SNF-Related, Matrix-associated, Actin-dependent Regulator of Chromatin, Subfamily A, Member 5), which functions cooperatively with *BAZ1A*[59]. By contrast, expression of the AR and VDR were unchanged between EA and AA PCa patients. Across the cell models, the expression and regulation of these proteins diverged between EA and AA cell models. *BAZ1A* expression was highest in LNCaP and unchanging in expression by $1\alpha,25(\text{OH})_2\text{D}_3$, whereas it was strongly induced in RC43N, but strongly repressed in RC43T. These patterns were broadly the same for *SMARCA5* (**Figure 6B**).

To identify the impact of altered *BAZ1A/SMARCA5* on expression of VDR target genes, we examined the correlations between either *BAZ1A* or *SMARCA5* and the $1\alpha,25(\text{OH})_2\text{D}_3$ -regulated genes from RC43N and RC43T in the AA *TPR52* fusion negative tumors from the TCGA-PRAD cohort. From these correlations we filtered those genes in Hallmarks_InflammatoryResponse and GO_Circadian Rhythm (**Supplementary Figure 11A**). Supportively, *BAZ1A* and *SMARCA5* correlations to these pathway genes were more pronounced for RC43N $1\alpha,25(\text{OH})_2\text{D}_3$ -regulated genes than those from RC43T, suggesting that *BAZ1A/SMARCA5* is a VDR co-activator and impacts expression of genes associated with inflammation. Next, we examined the genes in *BAZ1A* containing SWI/SNF complexes that were expressed in the cell lines and tumor cohorts. Expression patterns of the *BAZ1A* SWI/SNF complex genes significantly distinguished the AA from the EA cell line models (**Supplementary Figure 11B**). The most altered genes in this complex also significantly distinguished *TPR52* fusion positive from negative tumors in the TCGA-PRAD cohort (chi-squared $p = 0.002$), suggesting that genomic ancestry impacted expression of these genes was most common in the absence of *TPR52* fusion (**Supplementary Figure 11C**). Finally, we examined expression of all the genes in all four SWI/SNF complexes in the Rayford et al cohort of AA tumors ($n = 596$) compared to EA tumors ($n = 556$)[10]. All genes detected genes from each of the four complexes were significantly down-regulated in AA tumors compared to EA counterparts (**Supplementary Figure 11D**).

We therefore tested the impact of BAZ1A expression in the four cell line models using RNA-Seq (**Figure 6C, Supplementary Figure 8**). A BAZ1A-containing plasmid was transiently expressed in all four cells lines and elevated expression confirmed (**Supplementary Figure 12A**). In each model we undertook RNA-Seq after $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulation (100 nM, 8h) in BAZ1A over-expressed cells compared to vector controls. BAZ1A over-expressed cells displayed enhanced $1\alpha,25(\text{OH})_2\text{D}_3$ responses which was most pronounced in LNCaP and RC43N in terms of the number of genes with enhanced responses. Genes associated with a VDR ChIP-Seq peak or NF regions in the same cell background are represented as a red symbol with a border on the volcano plots (**Figure 6B**).

The VDR cistrome genes were in many cases significantly more enriched in than predicted by chance (**Table 7**). This supports a role for BAZ1A levels to augment regulation of $1\alpha,25(\text{OH})_2\text{D}_3$ -regulated genes. However, it is made more striking by the fact that the same analyses of the parental cells (**Figure 4A**) found no evidence for significant enrichment.

Cell	Cistrome	logPval	NumberCistromeGenes	Threshold	MostSignificant
RC43N	RC43N.VDR.ATAC	11.35	1358	Significant	PIK3R6
RC43T	RC43T.VDR.ATAC	3.01	349	Significant	PIK3C2G
RC43T	RC43T.VDR.ChIP	1.40	28	Significant	FRG2B
HPr1AR	HPr1AR.VDR.ATAC	0.82	73	NS	KCNH6
LNCaP	LNCaP.VDR.ATAC	0.40	80	NS	ANGPTL7
RC43N	RC43N.VDR.ChIP	0.28	14	NS	ANKRD26P1
LNCaP	LNCaP.VDR.ChIP	0.26	51	NS	POTEM

Table 7: Overlap of BAZ1A-dependent $1\alpha,25(\text{OH})_2\text{D}_3$ -regulated gene expression with genes annotated to $1\alpha,25(\text{OH})_2\text{D}_3$ -regulated nucleosome free regions and VDR binding sites significantly. Genes were annotated to ATAC-Seq or ChIP-Seq regions within 100kB and those genes overlapped with the differentially expressed genes in the same cell background was BAZ1A was overexpressed, and enrichment tested with a hypergeometric test (lower.tail = FALSE).

Analyses of the GSEA terms also supported a role for BAZ1A to impact gene expression. For example, in LNCaP cells the combined effect of BAZ1A/ $1\alpha,25(\text{OH})_2\text{D}_3$ elevated enrichment of genes associated with the Androgen response, UV response and decreased enrichment of MYC targets. By contrast, in RC43N the most enriched terms were Interferon γ and α , Inflammatory responses, IL6 signaling and TNFA signaling suggesting an immunomodulatory phenotype (**Figure 6 D**). Focusing on the impact of BAZ1A on $1\alpha,25(\text{OH})_2\text{D}_3$ -induced expression changes in RC43N and RC43T, and calculating the enrichment terms that change the most, in either the same or a divergent manner, revealed that BAZ1A has a potent and cell-specific impact on gene expression patterns induced by $1\alpha,25(\text{OH})_2\text{D}_3$ (**Figure 6 E**). For example, commonly enriched terms included many of the inflammation terms such as IL-6- and IL-2-mediated signaling, and TNFA signaling as well as Estrogen signaling and

apoptosis pathways, whereas divergent pathways included interferon signaling, MYC signaling and androgen responses.

VDR cistrome-transcriptome target genes in AA PCa clinical cohorts. Finally, we examined how these cistrome-transcriptome relationships established AA and EA cell lines could be detected in clinical cohorts of AA and EA men with PCa, using three cohorts. In the first instance, we examined serum samples from patients with high-grade prostatic intraepithelial neoplasia (HGPIN) who participated in a Southwest Oncology Group (SWOG) clinical trial (SWOG S9917) [60]. Within the cohort, 40% of men progressed from HGPIN to PCa and we examined miRNA that significantly predicted HGPIN progression to PCa in either EA or AA enrolled patients. In this manner we identified separate miRNA associated with HGPIN progression to PCa, and the top 20 in AA patients and EA patients are shown **Supplementary Table 6**. Twelve of the 33 miRNA (36%) that associated with AA progression were annotated to AA 1 α ,25(OH)₂D₃/VDR cistrome regions including *MIR23B* and *RTCA* (contains *MIR553*) (**Table 8**), whereas this was 37/280 (13%) for the EA progression miRNA.

Progression	Cistrome	NumberMIRNA	MostSignificant.MIRNA
AA_Prog	indep	21	MIR19A
AA_Prog	RC43T.VDR.ChIP	1	MIR23B
AA_Prog	RC43N.VDR.ATAC	4	MIR553
AA_Prog	RC43T.VDR.ATAC	7	MIR553
EA_Prog	indep	243	MIR412
EA_Prog	LNCaP.VDR.ChIP	2	MIR664A
EA_Prog	HPr1AR.VDR.ATAC	32	MIR2110
EA_Prog	LNCaP.VDR.ATAC	3	MIR3605

Table 8: Serum expression of miRNA that associate with progression from high-grade prostatic intraepithelial neoplasia to PCa in AA and EA patients. Serum samples from 96 patients with high-grade prostatic intraepithelial neoplasia (HGPIN) who participated in a Southwest Oncology Group (SWOG) clinical trial (SWOG S9917). The cohort consisted of 21 AA patients (9 progressed to PCa) and 75 EA patients (33 progressed to PCa). Nanostring PCR was used to measure differential miRNA expression within AA or EA samples and between EA and AA samples by progression status. The data were processed with NanoStringDiff and significantly different miRNA identified (logPV > 1 & absFC > .58). 33 miRNAs significantly associated with AA progression to PCa and more than 200 EA miRNA associated with PCa progression. The AA and EA miRNA genes (or host genes) that exclusively associated with progression were annotated to AA or EA 1 α ,25(OH)₂D₃/VDR cistrome regions, and the number of annotations is indicated and an example of the most significantly regulated miRNA given.

Next, we re-examined data from our earlier PCa RNA-Seq study from EA and AA patients who received vitamin D₃ (4000 IU daily) prior to radical prostatectomy. Ancestry informative markers confirmed the African genomic ancestry of the AA patients (data not shown). As we reported previously [7] the responses in EA patients were essentially null, whereas a strong transcriptional response was

observed in the AA patients with ~1400 genes being significantly regulated by vitamin D₃, and GSEA revealed these genes were enriched in immuno-modulatory and PCa-relevant pathways (**Figure 7 A, B**). Remarkably, 326 (~23%) of these genes were significantly bound by VDR in ChIP-Seq studies. The DEGs were again enriched for inflammatory signaling components. Interestingly, 70% of the significantly modulated genes in the AA patients were either VDR ChIP-Seq or ATAC-Seq annotated genes (**Table 9**)

AA.ChIP.targets	AA.ATAC.targets	NumberTargets	Percent
AA.ChIP	AA.ATAC	55	70.5
AA.ChIP	indep	23	29.5
indep	AA.ATAC	950	69.2
indep	indep	423	30.8

Table 9: Percentage overlap of vitamin D₃-regulated gene expression in prostate tumors from AA patients with genes annotated by 1 α ,25(OH)₂D₃-regulated nucleosome free regions and/or VDR ChIP-Seq. A cohort of seven AA PCa patients with conserved African genomic ancestry and 16 EA PCa patients were treated with vitamin D₃ (4000 IU daily), and RNA-Seq undertaken on the tumors following radical prostatectomy, as we reported previously[7]. Significantly differentially regulated genes in the AA PCa group (there were no DEGs in the EA PCa group) were overlapped with genes annotated to ATAC-Seq or ChIP-Seq regions within 100kB. The percentage overlap of DEGs and cistrome genes is shown.

Finally, in a further cohort of 57 AA and 18 EA patients we undertook RNA-Seq in tumor and contralateral normal material collected from clinics in Chicago, IL [61]. DEG analyses were undertaken in tumors and normal material in two comparisons to identify the race-specific impact of serum 25(OH)D₃ levels and obesity. Firstly, we compared how gene expression differed between in the cohort with deficient serum 25(OH)D₃ levels (< 12 ng/ml) compared to replete levels, while keeping obesity as a continuous variable. Secondly, we compared how gene expression differed between the obese and normal groups, while keeping serum 25(OH)D₃ levels as a continuous variable.

In the AA cohort there were 1612 DEGs in the tumors with low serum 25(OH)D₃ levels compared to tumors from patients with replete levels. Reflecting, the radical prostatectomy samples, there were no significant DEGs in EA patients when considering the levels of serum 25(OH)D₃. Similarly, in the AA cohort comparing gene expression in tumors from men who are obese compared to normal counterparts there were 2415 DEGs; again, there were no DEGs in the EA tumor group between obese and normal BMI patients. Finally, these DEGs were queried as to whether they were also annotated to be VDR ChIP-Seq or 1 α ,25(OH)₂D₃ stimulated NF ATAC-Seq genes (**Figure 7C**). **Table 10** summarizes the overlaps of these DEGs and cistrome data with 85% of either DEG group (impact of serum 25(OH)D₃ or obesity) overlapping with the AA cistrome data.

BMI	D3	AA.ChIP.targets	AA.ATAC.targets	NumberTargets	Percent
adj	low	AA.ChIP	AA.ATAC	78	85.7
adj	low	AA.ChIP	indep	13	14.3
adj	low	indep	AA.ATAC	1217	80.0
adj	low	indep	indep	304	20.0
O	adj	AA.ChIP	AA.ATAC	105	85.4
O	adj	AA.ChIP	indep	18	14.6
O	adj	indep	AA.ATAC	1852	80.8
O	adj	indep	indep	440	19.2

Table 10: Percentage overlap of either serum 25(OH)D₃ vitamin D₃-regulated or obesity-regulated gene expression in prostate tumors from AA patients with genes annotated by 1 α ,25(OH)₂D₃-regulated nucleosome free regions and/or VDR ChIP-Seq. RNA-Seq was undertaken in tumors from a cohort of 57 AA and 18 EA patients who underwent radical prostatectomy. Tumor-specific significant differentially expressed genes (DEGs) were identified for deficient serum 25(OH)D₃ (serum 25(OH)D₃ levels < 12 ng/ml; low) or obesity (BMI > 30; O). In each case BMI or 25(OH)D₃ levels were kept as a continuous variable respectively (adj). The DEGs in the AA PCa group (there were no DEGs in the EA PCa group) were overlapped with genes annotated to ATAC-Seq or ChIP-Seq regions within 100kB. The percentage overlap of DEGs and cistrome genes is shown.

DISCUSSION

The current study aimed to define if $1\alpha,25(\text{OH})_2\text{D}_3$ and VDR signaling capacity could provide insights about biological mechanisms involved in racial disparities in PCa between African American and European American men. Strong epidemiological findings, as well as clinical trial data broadly support the concept that VDR signaling activated by $1\alpha,25(\text{OH})_2\text{D}_3$ exerts a tumor suppressive effect in AA PCa patients, and potentially that this patient group is also most acutely vulnerable to a lack of $1\alpha,25(\text{OH})_2\text{D}_3$ activated VDR signaling. However, genomic insight into these relationships is not well established, and our proteogenomic study sought to fill this knowledge gap.

To define VDR actions across AA and EA prostate models we generated and integrated high dimensional genomic, transcriptomic and proteomic data sets, combined with clinical genomic data to allow a comprehensive insight into how VDR signaling is altered within and across races. Our findings support several key insights and inferences on VDR signaling, and perhaps are also relevant to a larger set of nuclear receptors in the context of health disparities. Furthermore, to the best of our knowledge, this study represents the first comparative cistromic study between AA and EA cell models and these findings may be insightful to other workers addressing cancer and race-related health disparities.

A starting point for our study was to consider expression of the VDR protein and capture its interactome, which established that while prominent VDR expression is shared in the EA and AA cell models, the interacting proteins and cellular $1\alpha,25(\text{OH})_2\text{D}_3$ sensitivity varies widely. For example, the VDR complex differed by how it was associated with coactivators such as SMARCA5, splicing factors such as DDX39B, and the transcriptional regulator TRIM29. Beyond this, proteins that are also part of the regulation of circadian rhythm were enriched in the VDR complex, such as a gain of ARNTL2 and loss of NONO in RC43T cells, supporting a role for VDR to cooperate with the control of circadian rhythms.

Next, we considered the genomic impact of $1\alpha,25(\text{OH})_2\text{D}_3$ by examining the overlap in nucleosome positioning and where the VDR binds. These approaches established that $1\alpha,25(\text{OH})_2\text{D}_3$'s genomic impact is greatest in RC43N and RC43T, but divergent between these cells. $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated NF regions in RC43N occurred more frequently than in other models, were largely unique and >98% represented a gain. In contrast, $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated NF regions in RC43T were less frequent and resulted in an overall NF region loss. This suggests that VDR is primed to produce very different responses between these two cells, although motif enrichment suggested that the TFs that

are enriched in these regions are similar and include AP-1 factors, such as KLF5. The divergence of this response was illustrated by changes in enrichment of circadian rhythm TF motifs, such as CLOCK.

Similarly, ChIP-Seq revealed strong agreement for where the VDR bound in the absence of $1\alpha,25(\text{OH})_2\text{D}_3$ across LNCaP, RC43N and RC43T with more than 300 shared binding sites. The addition of $1\alpha,25(\text{OH})_2\text{D}_3$ resulted in distinct patterns of VDR binding within a cell, and reduced the number of shared sites with only nine sites in common across all cells. Furthermore, reflecting the NF data, there was a dramatic reduction in VDR binding in RC43T. Motif enrichment supported the concept that the basal VDR state in LNCaP and RC43T was the most TF active state with a clear loss of motif enrichment when $1\alpha,25(\text{OH})_2\text{D}_3$ was present. In RC43T, the loss of VDR binding reflected the loss of NF regions and again suggests that the system is altered to limit the impact of $1\alpha,25(\text{OH})_2\text{D}_3$. Reflecting the genomic data, at the transcriptional level the RC43N and RC43T cells were most responsive and indeed many genes that were modulated by $1\alpha,25(\text{OH})_2\text{D}_3$ also were either bound by VDR, or had a change in $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated NF distribution. Again, the response between RC43N and RC43T were frequently opposed, with many GSEA terms being switched in direction between the cells.

Together these data support the concept that RC43N and RC43T have the strongest response to $1\alpha,25(\text{OH})_2\text{D}_3$ but the effect is highly divergent between them. We therefore sought mechanisms that may drive this divergence. Analyses of TCGA data identified a potential role for altered expression of the BAZ1A/SMARCA5 SWI/SNF complex. Therefore, we tested this possibility. Again, underscoring the differences between RC43N and RC43T, $1\alpha,25(\text{OH})_2\text{D}_3$ -treatment upregulated BAZ1A and SMARCA5 in RC43N, but down-regulated these proteins in RC43T. It is noteworthy that SMARCA5 was significantly less enriched in the VDR complex in RC43T compared to LNCaP, that RC43N VDR ChIP-Seq significantly overlapped with publicly available SMARCA4 ChIP-Seq from LNCaP cells, and that $1\alpha,25(\text{OH})_2\text{D}_3$ treatment resulted in isoform-specific expression of SMARCA4 in RC43N. It is also worth noting that GWAS SNPs in BAZ1A associate significantly with heel bone strength, also supporting a role in regulating VDR function [62]. Restoring BAZ1A expression led to significantly enhanced $1\alpha,25(\text{OH})_2\text{D}_3$ -regulated transcriptome, but these genes were only significantly enriched for VDR bound and NF regions in RC43N and RC43T suggesting that BAZ1A expression had the most significant impact on VDR function in AA cells. The most strongly up-regulated gene in RC43N that is annotated to RC43N $1\alpha,25(\text{OH})_2\text{D}_3$ -regulated NF region was *PIK3R6*, which recently was identified as a novel antigen in a clinical immunotherapy trial in advanced PCa[63]. These data also contribute to the earlier

analyses of VDR functions in the prostate[64] and also mechanisms by which VDR limits proliferation[65].

By identifying AA and EA specific VDR genomic events, we measured if these differences were apparent in three clinical cohorts. These analyses revealed that the footprint of VDR signaling was most apparent in AA PCa. For example, miRNA that predicted progression from HGPIN to PCa in AA men were highly enriched for VDR cistrome data, as were genes that responded in prostate tumors from men receiving vitamin D3 supplementation prior to radical prostatectomy. This was also strikingly apparent in prostate tumors from men who had deficient serum 25(OH)D₃ levels, and indeed in this interacted significantly with obesity (BMI >30.0 kg/m²) status.

It has been clear for many years [66] that the vitamin D signaling axis has systemic functions beyond the control of calcium regulation. Indeed, large scale transcriptomic and proteomic studies have revealed an interesting paradox, namely that the VDR is expressed widely and robustly across cell-types including those that have no obvious role in calcium function [67, 68]. Furthermore the impact of altered vitamin D levels is strongly implicated in the etiology of non-calcemic syndromes such as diabetes cardiovascular disease, autoimmune disease, and various high-profile cancers. Recently, in the VITAL cohort, vitamin D3 has been shown to reduce the incidence of autoimmune diseases [69].

Although guidance is available for serum 25(OH)D₃ levels required for bone health, it is far from clear what level is required either to promote cardiovascular health or to prevent autoimmune diseases. It is even less clear how vitamin D deficiency amongst African Americans, which is highly prevalent, impacts cancer and these other diseases. Our genomic data suggest that the powerful example of changing melanin content in the skin through adaptation has occurred in parallel with a range of cellular mechanisms that include the genomic functions of the VDR in non-calcemic tissues.

Consistently, through this study, the data support a model whereby VDR signaling is more potent in AA prostate cell models and prostate tumors. VDR-induced enrichment of circadian rhythm and inflammatory networks are evident. Observational evidence supports a cross-talk between VDR and components of the transcriptional network that regulates melatonin production, circadian rhythm and sleep duration[40, 70]. Given the manner in which VDR function can be corrupted, either through changes in serum Vitamin D3 levels[15, 16, 30, 71, 72], changes in membrane transport[27] or disruption to the composition of the VDR complex [73-81], it is reasonable to infer that the VDR stands at the crossroads of biopsychosocial signaling that impacts PCa by a distinct pattern of VDR genomic

binding in cells with African genomic ancestry. That these VDR patterns in prostate cells are distinctly enriched for genes that control of circadian rhythm and inflammatory networks suggest that men of African ancestry may be disproportionately impacted by environments of low UVB exposure, increase stress, access to quality early detection and treatment, as well work environments that are deleterious for circadian signaling. To this concept we have added the mechanistic insight that the SWI/SNF complexes containing BAZ1A and SMARCA5 are distorted in a manner that reflects genomic ancestry, and further distorts the normal functions of the VDR and suggests enhanced sensitivity to DNA damage[82].

In conclusion, the current study suggests that the functions of the VDR may have adapted with significant distinctions between people of different genomic ancestry. More specifically, we reason that adaptation to environments of lower UVB exposure that could potentially be associated with vitamin D insufficiency and alter the prominence of how VDR signaling occurs in a wide variety of tissues. Furthermore, we propose that the prostate is an important gland with which to test this possibility given it is the site of significant syndromes and diseases that are highly impactful on US men, and furthermore many of these conditions display significant health disparities.

MATERIALS AND METHODS

Cell culture and materials. The EA cells HPr1AR, LNCaP (140) and the AA cells RC43N and RC43T were maintained at 37°C and 5.0% CO₂ using a cell culture incubator with UV contamination control (Sanyo). HPr1AR cells were maintained in keratinocyte serum free media (supplemented with 25mg of Bovine pituitary extract, 10ug epithelial growth factor and 10% FBS); LNCaP cells in RPMI 1640 Medium containing 10% FBS; RC42N and RC43T in keratinocyte serum free media (supplemented with 25mg of Bovine pituitary extract, 10ug epithelial growth factor and 10% FBS). All media was supplemented with 100 U/mL Penicillin-Streptomycin. 1 α ,25(OH)₂D₃ was kept as 10mM EtOH stocks, and diluted to 1000x stocks prior to treatments. Cell lines were authenticated by STR profiling and confirmed mycoplasma free by RT-PCR.

RT-qPCR. Total RNA was isolated via TRIzol® reagent (Thermo Fisher Scientific) for candidate mRNA detection by use of the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen), following manufacturer's protocols. Complementary DNA (cDNA) was prepared using iScript™ cDNA Synthesis Kit (Bio-Rad), following manufacturer's protocols. Relative gene expression was subsequently quantified via Applied Biosystems 7300 Real-Time PCR System (Applied Biosystems), for both TaqMan and SYBR Green (Thermo Fisher Scientific) applications. All targets were detected using either pre-designed TaqMan Gene Expression Assays (Thermo Fisher Scientific; *VDR*, *BAZ1A*, *SMARCA5* using pre-designed PrimeTime qPCR primers and a final primer concentration of 500nM. All primers for use with SYBR Green application were tested for specificity by melting curve analysis with subsequent product visualization on agarose gel. All RT-qPCR experiments were performed in biological triplicates, with at least technical duplicates. Fold changes were determined using the 2^{-ΔΔCt} method as the difference between experimental group and respective control group. Significance of experimental comparisons was performed using Student's t-test.

Stable transfection of BAZ1A. GFP-BAZ1A was purchase from Addgene (plasmid # 65371)[83] and transfection was cells were transfected and stable selection after transduction with by selection and maintainance in media supplemented with puromycin (2μg/mL).

Western Immunoblotting. Total cellular protein was isolated from exponentially growing cells for determination of target protein expression. Cells were harvested, then washed in ice cold PBS before lysing in ice cold RIPA buffer (50mM Tris-HCl pH 7.4, 150mM NaCl, 1% v/v Triton X-100, 1mM EDTA pH 8.0, 0.5% w/v sodium deoxychlorate, 0.1% w/v SDS) containing 1x cOMplete Mini Protease Inhibitor

Tablets (Roche). Protein concentrations were quantified using the DC Protein Assay (Bio-Rad), following manufacturer's protocols. Equal amounts of proteins (30-60µg) were separated via SDS polyacrylamide gel electrophoresis (SDS-PAGE) using precast 4-20% gradient polyacrylamide gels (Mini-Protean TGX, Bio-Rad). Proteins were transferred onto polyvinylidene fluoride (PVDF) membrane (Roche) for 30V for 16 hours. Post transfer, membranes were blocked with 5% non-fat dry milk (NFDM) for 1 hour at room temperature. Blocked membranes were probed with primary antibody against BAZ1A, SMARCA5, VDR, GAPDH either overnight at 4°C or for 3 hours at room temperature. Primary antibody was detected after probing for 1 hour with HRP-linked rabbit anti-mouse IgG (P0161, Dako) or goat anti-rabbit IgG (P0448, Dako) secondary antibody at room temperature using ECL Western Blotting substrate (Pierce). Signal quantification was performed using the ProteinSimple Fluorochem M Imager.

Cell viability. Bioluminescent detection of cellular ATP as a measure of cell viability was undertaken using CellTiter-Glo® (promega) reagents. Cells were plated at optimal seeding density to ensure exponential growth in 96-well, white-walled plates. Wells were dosed with agents to a final volume of 100 µl. Dosing occurred at the beginning of the experiment, and cells were incubated for up to 120 hours. Luminescence was detected with Synergy™ 2 multi-mode microplate reader (BioTek® Instruments). Each experiment was performed in at least triplicate wells in triplicate experiments.

Clonogenic assays. The colony formation ability of cells in presence of variable doses of 1α,25(OH)₂D₃. 1000 cells were plated in triplicates in a 6 well plate and treated with 1α,25(OH)₂D₃ every three days for a period of 14days. After 14 days cells were washed and fixed with neutral buffered formalin and stained with crystal violet stain and quantified as previously reported[84].

Rapid immunoprecipitation mass spectrometry of endogenous protein. RIME analyses were undertaken with antibody towards the VDR in cells treated with either vehicle or 1α,25(OH)₂D₃. 20x10⁶ cells were crosslinked with 1% formaldehyde solution, quenched with glycine (0.1 M, pH 7.5) and harvested in cold PBS. Nuclei were separated as previously reported [85] and subjected to sonication (30s on 30s off cycles for 30mins) for genomic DNA fragmentation. 30ul of 10% triton-X was added and high-speed centrifugation was performed to separate the nuclear proteins. Further, VDR and IgG antibody conjugated beads were incubated with nuclear lysates overnight and washed ten times with RIPA buffer and 2 washes of Ambic solution as previously described [85]. LC-MS/MS was performed over a 2-hour separation and mean spectral count results were analyzed with a generalized linear model workflow to identify differentially enriched proteins.

RNA-Seq. RNA was extracted from cells in the presence of $1\alpha,25(\text{OH})_2\text{D}_3$ (100nM, 8hr) or EtOH in biological triplicate samples and analyzed by RNA-Seq. Sequencing was performed at the Nationwide Children's Hospital Institute for Genomic Medicine, and sequencing libraries prepared with the TruSeq Stranded Total RNA kit (Illumina Inc), from 1ug total RNA. Alignment of raw sequence reads to the human transcriptome (hg38) was performed via `Rsubread` [86] and transcript abundance estimates were normalized and differentially expressed genes (DEGs) identified using a standard `edgeR` pipeline. Functional annotation of gene sets: Pathway enrichment analysis and gene set enrichment analysis (GSEA) were performed using gene sets from the Molecular signatures database (MSigDB).

Small RNA-Seq. Cell lines were treated as RNA-Seq and sequencing performed at Nationwide Children's Hospital Institute for Genomic Medicine. Sequencing libraries were prepared with the TruSeq Small RNA kit (Illumina Inc), from 1ug total RNA. Following manufacturer's protocols, ligation of 5' and 3' RNA adapters to the mature miRNAs 5'-phosphate and 3'-hydroxyl groups, respectively was undertaken. Following cDNA synthesis, the cDNA was amplified with 11-13 PCR cycles using a universal primer and a primer containing one of 48 index sequences, which allowed pooling of libraries and multiplex sequencing. Prior to pooling, each individual sample's amplified cDNA construct was visualized on a DNA-HS Bioanalyzer DNA chip (Agilent Technologies) for mature miRNA and other small RNA products (140-150bp). Successful constructs were purified using a Pippen prep (Sage Inc.), using 125 – 160 bp product size settings with separation on a 3% agarose gel. The purified samples were validated for size, purity and concentration using a DNA-HS Bioanalyzer chip. Validated libraries were pooled at equal molar to a final concentration of 10nM in Tris-HCl 10 mM, pH 8.5, before 50 cycle sequencing on a MiSeq (Illumina, Inc.). Fastq files were aligned to the genome (hg38) using `Rsubread`. Expression counts were called against the miRbase consensus miRnome using `featureCounts` and A standard `edgeR` pipeline determined differentially expressed miRNA.

ChIP-Seq. ChIP was performed cells in the presence of $1\alpha,25(\text{OH})_2\text{D}_3$ 2D_3 (100nM, 6hr) or EtOH in triplicate or EtOH in triplicate independent experiments. Briefly, approximately 20×10^6 cells were crosslinked with 1% formaldehyde solution, quenched with glycine (0.125 M) and harvested in cold PBS. Sonication of crosslinked chromatin was performed using a Bioruptor® UCD-200™ Sonicator (Diagenode) with optimized cycles for each cell type. Immunoprecipitation of sonicated material was performed with antibodies against VDR (D2K6W) – Cell Signaling) or IgG (Rabbit IgG sc2729 – Santa Cruz) for 16 hours, and antibody/bead complexes isolated with Magna ChIPTM Protein A+G magnetic beads (Millipore). Complexes were washed, reverse crosslinked, and treated sequentially with RNase

and proteinase K prior to DNA isolation. Sequencing was performed at the Nationwide Children's Hospital Institute for Genomic Medicine. The VDR cistrome was analyzed with `Rsubread/csaw` [87] along with TF motif analyses (`MotifDb`).

In order to find potential transcription factor binding enrichment within VDR cistromes, `GIGGLE` was utilized to query the complete human transcription factor ChIP-seq dataset collection (10,361 and 10,031 datasets across 1,111 transcription factors and 75 histone marks, respectively) in Cistrome DB[51]. Prostate specific filtering limited analysis to 681 datasets across 74 TFs and 238 datasets across 19 HMs. For each query dataset, we determined the overlap of each VDR cistrome. Putative co-enriched factors were identified by assessment of the number of time a given factor was observed in the top 200 most enriched datasets relative to the total number of datasets for that factor in the complete Cistrome DB (> 1.2 FC enrichment over background). For prostate specific analysis, overlaps across datasets were averaged for each factor.

ATAC-Seq. 50000 cells were plated and treated with $1\alpha,25(\text{OH})_2\text{D}_3$ and vehicle control for 4h and cells were washed twice with PBS and collected by trypsinization. Cells were resuspended in 50ul of ATAC-resuspension buffer (ATAC-RSB - 10mM Tris-HCl, 10mM NaCl, 3mM MgCl_2) containing (0.1% NP-40, 0.1% tween-20, and 0.01% digitonin) and pipetted up and down 3 times. Further, 1ml of ATAC-wash-resuspension buffer (ATAC-RSB + 0.1% tween 20) was used to pellet down the nuclei. The nucleic were further resuspended in transposition mix (2X TD buffer, 1X PBS, Digitonin 0.01%, tween 20 0.1%, NFW 5ul, and Illumina transposase 2.5ul). Mixing, cleanup and library preparation, quantification and sequencing was performed using NovaSeq6000 S1 PE150bp Sequencing as per protocol (28846090). (ATAC-Seq data were separated into nucleosome free (NF), mono-, di- and tri-nucleosome compartments (`ATACSeqQC`)[88].

MiRNA expression in serum samples. In serum samples from SWOG S9917 (high-grade prostatic intraepithelial neoplasia (HGPIN) to PCa), Nanostring was used to identify miRNA associated with progression in AA patients. Nanostring PCR was used to measure differential miRNA expression within race and across race by progression status. The data were processed with `NanoStringDiff` and significantly different miRNA identified ($\log\text{PV} > 1$ & $\text{absFC} > .58$).

Data analyses and integration. All analyses were undertaken using the R platform for statistical computing (version 3.6.1) and the indicated library packages implemented in Bioconductor.

Declarations

Ethics approval and consent to participate

The serum samples from the Southwest Oncology Group (SWOG) clinical trial (SWOG S9917) were collected under local IRB approval [60]. The PCa samples from EA and AA patients who received vitamin D₃ prior to radical prostatectomy were collected under local IRB approval [7]. The radical prostatectomy PCa samples from EA and AA patients prior were collected under local IRB approval[61]

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analysed during the current study will be available on GEO

Competing interests

C.Y is consultant and shareholder for Riptide Biosciences. CY has received honorarium and consultant fees from QED Therapeutics and Amgen. The remaining authors certify that he has NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Funding

MJC, *CY*, *ABM*, and *LESC* acknowledge support in part from the Prostate program of the Department of Defense Congressionally Directed Medical Research Programs [W81XWH-20-1-0373; W81XWH-21-1-0850], from U54-MD007585-26 (NIH/NIMHD), U54 CA118623 (NIH/NCI), and Department of Defense Grant (PC170315P1, W81XWH-18-1-0589) awarded to *CY*. *CHH* and *GH* were supported by grant #U54MD010706 from the National Institute on Minority Health and Health Disparities. *MJC* also acknowledges National Institute of Health Cancer Center Support Grant (P30CA016058) to the OSUCCC The James. *CHH* and *GH* were supported by grant #U54MD010706 from the National Institute on Minority Health and Health Disparities.

Authors' contributions

MS undertook ChIP-Seq, RNA-Seq and cell biology; *SS* undertook ATAC-Seq, RIME, cell biology; *SAW* contributed to ChIP-Seq, ATAC-Seq and RIME; *HT* contributed to bioinformatic analyses; *JSG* contributed to ATAC-Seq and cell biology; *HJ*, *HW* contributed to cell biology; *MDL* undertook GIGGLE analyses; *IE* undertook ancestry informative analyses; *BK*, *HW* contributed to experimental design; *RM*, *GH* undertook the RNA-Seq of the vitamin D supplementation trial in prostate cancer patients; *IBA*, *SOR* contributed to experimental design and conceptual framework; *ARM*, *MBD* and *RAK* designed and oversaw the RNA-Seq of the tumors from prostate cancer patients and captured the clinical data including serum 25(OH)D₃ levels; *LN* contributed to data interpretation; *JRM* was a co-lead on SWOG S9917 and supplied serum samples for analyses; *CHH* and *GH* oversaw and designed the vitamin D supplementation trial in prostate cancer; *LES-C*, *CY* and *MJC* jointly conceived of the study design; *MJC* oversaw the implementation of the study and undertook other bioinformatic analyses and generated tables and figures.

Figure legends

Figure 1. Expression of VDR and responses to $1\alpha,25(\text{OH})_2\text{D}_3$ in African and European American Cell lines. **A.** The European American (EA) non-malignant HPr1AR and malignant LNCaP prostate cells and the African American (AA) non-malignant RC43N and RC77N, and malignant RC43T and RC77T cells were plated at optimal seeding density in 96 well white walled plates and grown for 96h with the indicated nM doses of $1\alpha,25(\text{OH})_2\text{D}_3$ or vehicle control. Measurements of cellular levels of ATP, as an indicator of cell viability were performed. Each measurement was performed in biological triplicates and in triplicate wells. **B.** Clonal proliferation was measured after 10 days growth with the indicated doses of $1\alpha,25(\text{OH})_2\text{D}_3$ or vehicle control. **C.** Western immunoblot measurements of VDR levels after $1\alpha,25(\text{OH})_2\text{D}_3$ treatment (100 nM, 24h) or vehicle control. **D.** RIME analyses was undertaken for VDR in the indicated cells and significantly different proteins were identified using an edgeR workflow. Volcano plots depicting expression levels compared to IgG controls in basal or $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated conditions (100 nM, 4h). Significant ($p.\text{adj} < .1$) differentially and uniquely enriched proteins in each cell and treatment condition were classified either as a Coactivator (CoA), Corepressor (CoR), Mixed function coregulator (Mixed) or transcription factor (TF). **E.** Circadian rhythm regulators were selected and visualized with bars with a yellow border being significant.

Figure 2. VDR ATAC-Seq in African and European American Cell lines. **A.** ATAC-Seq was undertaken in triplicate in HPr1AR, LNCaP, RC43N and RC43T following $1\alpha,25(\text{OH})_2\text{D}_3$ treatment (100 nM, 4h) or vehicle control. FASTQ files were QC processed, aligned to hg38 (Rsubread), sorted and duplicates removed before further processing with ATACseqQC to generate nucleosome free and mono-nucleosome regions. Differential enrichment of regions was measured with csaw and the significantly different regions ($p.\text{adj} < .1$) were then intersected to generate the Venn diagrams of overlapping regions by a minimum of 1bp (ChIPpeakAnno). **B.** Motif enrichment in NF regions was undertaken with Homer and ranked by significance to visualize in descending significance. **C.** Frequency of enriched motifs by transcription factor families' class. **D.** Enrichment of circadian rhythm transcription factors.

Figure 3. VDR ChIP-Seq in African and European American Cell lines. **A.** Basal and $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated (100 nM, 6h) VDR ChIP-Seq was undertaken in triplicate in HPr1AR, LNCaP, RC43N and RC43T. FASTQ files were QC processed, aligned to hg38 (Rsubread), sorted and duplicates removed before differential enrichment of regions was measured with csaw and the significantly different regions compared to IgG controls ($p.\text{adj} < .1$) were then intersected to generate the Venn diagrams of overlapping regions by a minimum of 1bp (ChIPpeakAnno). **B.** Significantly differentially enriched regions were identified by Homer and nuclear receptors are illustrated. **C.** Changes in motif enrichment were calculated (delta) and ranked by significance for visualization.

Figure 4. $1\alpha,25(\text{OH})_2\text{D}_3$ -dependent RNA- and small RNA-Seq in African and European American Cell lines. **A. B.** Basal and $1\alpha,25(\text{OH})_2\text{D}_3$ -stimulated (100 nM, 8h) RNA-Seq was undertaken in triplicate in HPr1AR, LNCaP, RC43N and RC43T. FASTQ files were QC processed, aligned to hg38 (Rsubread), and processed with limma-voom and edgeR workflow to identify significant differentially enriched genes (DEGs) ($\log\text{PV} > 1$ & $\text{absFC} > .37$) are illustrated on Volcano plots with top DEGs illustrated. **C.** GSEA analyses (Chemical Perturbations in GSEA) was undertaken and the terms identified and visualized where the normalized enrichment score was in the opposite direction in RC43N and RC43T.

Figure 5. VDR cistrome-transcriptome relationships. **A.** VDR binding sites from the ChIP-Seq data were annotated to DEGs within 100 kb and the sum of peak scores combined with gene expression magnitude were compared across cells, and classified according to ChromHMM distributions. **B.** As above but for ATAC-Seq data.

Figure 6. The impact of BAZ1A on expression of VDR-dependent gene networks. **A.** Altered *BAZ1A* and *SMARCA5* expression in African American PCa was identified in the TCGA prostate cancer cohort by comparing EA and AA tumors, and considering status of *TPR52* translocations. **B.** Western immunoblot measurements of *BAZ1A* and *SMARCA5* levels after $1\alpha,25(\text{OH})_2\text{D}_3$ treatment (100 nM, 24h) or vehicle control in cell lines. **C.** Unique DEGs in cells with restored expression of *BAZ1A*. **D.** *BAZ1A*-dependent DEGs enhanced inflammatory responses and repressed *MYC* networks. **E.** Comparable and divergent GSEA enrichment in *BAZ1A*-dependent DEGS in RC43N and RC43T

Figure 7. VDR cistrome-transcriptome relationships in prostate cancer. **A.** A cohort of seven AA PCa patients with conserved African genomic ancestry and 16 EA PCa patients were treated with vitamin D_3 (4000 IU daily), and RNA-Seq undertaken on the tumors following radical prostatectomy, as we reported previously[7]. Significantly differentially regulated genes in the AA PCa group (there were no DEGs in the EA PCa group) were overlapped with genes annotated to ATAC-Seq or ChIP-Seq regions within 100kb. The Volcano plot of the DEGs for the response in AA men, or comparing basal AA to EA PCa and annotated with genes that are VDR bound and/or $1\alpha,25(\text{OH})_2\text{D}_3$ -dependent NF region annotated genes. **B.** GSEA analyses of the vitamin D_3 stimulated AA tumors **C.** RNA-Seq was undertaken in tumors from a cohort of 57 AA and 18 EA patients who underwent radical prostatectomy. Tumor-specific significant DEGs were identified for deficient serum $25(\text{OH})\text{D}_3$ (serum $25(\text{OH})\text{D}_3$ levels < 12 ng/ml; low) or obesity (BMI > 30; O). In each case BMI or $25(\text{OH})\text{D}_3$ levels were kept as a continuous variable respectively. The DEGs for $25(\text{OH})\text{D}_3$ deficiency (left) of obesity (right) in the AA PCa group (there were no DEGs in the EA PCa group) were overlapped with genes annotated to ATAC-Seq or ChIP-Seq regions within 100kb.

References.

1. Patel HD, Doshi CP, Koehne EL, Hart S, Van Kuiken M, Quek ML, Flanigan RC, Gupta GN: **African American Men have Increased Risk of Prostate Cancer Detection Despite Similar Rates of Anterior Prostatic Lesions and PI-RADS Grade on Multiparametric Magnetic Resonance Imaging.** *Urology* 2021.
2. Fiorica PN, Schubert R, Morris JD, Abdul Sami M, Wheeler HE: **Multi-ethnic transcriptome-wide association study of prostate cancer.** *PLoS One* 2020, **15**:e0236209.
3. Hoffmann TJ, Van Den Eeden SK, Sakoda LC, Jorgenson E, Habel LA, Graff RE, Passarelli MN, Cario CL, Emami NC, Chao CR, et al: **A large multiethnic genome-wide association study of prostate cancer identifies novel risk variants and substantial ethnic differences.** *Cancer Discov* 2015, **5**:878-891.
4. Helfand BT, Roehl KA, Cooper PR, McGuire BB, Fitzgerald LM, Cancel-Tassin G, Cornu JN, Bauer S, Van Blarigan EL, Chen X, et al: **Associations of prostate cancer risk variants with disease aggressiveness: results of the NCI-SPORE Genetics Working Group analysis of 18,343 cases.** *Hum Genet* 2015, **134**:439-450.
5. Awasthi S, Berglund A, Abraham-Miranda J, Rounbehler RJ, Kensler K, Serna A, Vidal A, You S, Freeman MR, Davicioni E, et al: **Comparative Genomics Reveals Distinct Immune-oncologic Pathways in African American Men with Prostate Cancer.** *Clin Cancer Res* 2021, **27**:320-329.
6. Yuan J, Kensler KH, Hu Z, Zhang Y, Zhang T, Jiang J, Xu M, Pan Y, Long M, Montone KT, et al: **Integrative comparison of the genomic and transcriptomic landscape between prostate cancer patients of predominantly African or European genetic ancestry.** *PLoS Genet* 2020, **16**:e1008641.
7. Hardiman G, Savage SJ, Hazard ES, Wilson RC, Courtney SM, Smith MT, Hollis BW, Halbert CH, Gattoni-Celli S: **Systems analysis of the prostate transcriptome in African-American men compared with European-American men.** *Pharmacogenomics* 2016, **17**:1129-1143.
8. Yamoah K, Asamoah FA, Abrahams AOD, Awasthi S, Mensah JE, Dhillon J, Mahal BA, Gueye SM, Jalloh M, Farahani SJ, et al: **Prostate tumors of native men from West Africa show biologically distinct pathways-A comparative genomic study.** *Prostate* 2021, **81**:1402-1410.
9. Koga Y, Song H, Chalmers ZR, Newberg J, Kim E, Carrot-Zhang J, Piou D, Polak P, Abdulkadir SA, Ziv E, et al: **Genomic Profiling of Prostate Cancers from Men with African and European Ancestry.** *Clin Cancer Res* 2020, **26**:4651-4660.
10. Rayford W, Beksac AT, Alger J, Alshalalfa M, Ahmed M, Khan I, Falagario UG, Liu Y, Davicioni E, Spratt DE, et al: **Comparative analysis of 1152 African-American and European-American men with prostate cancer identifies distinct genomic and immunological differences.** *Commun Biol* 2021, **4**:670.
11. Fitzgerald LM, Agalliu I, Johnson K, Miller MA, Kwon EM, Hurtado-Coll A, Fazli L, Rajput AB, Gleave ME, Cox ME, et al: **Association of TMPRSS2-ERG gene fusion with clinical characteristics and outcomes: results from a population-based study of prostate cancer.** *BMC Cancer* 2008, **8**:230.
12. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, et al: **Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer.** *Science* 2005, **310**:644-648.
13. Liu W, Zheng SL, Na R, Wei L, Sun J, Gallagher J, Wei J, Resurreccion WK, Ernst S, Sfanos KS, et al: **Distinct Genomic Alterations in Prostate Tumors Derived from African American Men.** *Mol Cancer Res* 2020, **18**:1815-1824.
14. Magi-Galluzzi C, Tsusuki T, Elson P, Simmerman K, LaFargue C, Esgueva R, Klein E, Rubin MA, Zhou M: **TMPRSS2-ERG gene fusion prevalence and class are significantly different in prostate cancer of Caucasian, African-American and Japanese patients.** *Prostate* 2011, **71**:489-497.
15. Layne TM, Weinstein SJ, Graubard BI, Ma X, Mayne ST, Albanes D: **Serum 25-hydroxyvitamin D, vitamin D binding protein, and prostate cancer risk in black men.** *Cancer* 2017, **123**:2698-2704.

16. Nelson SM, Batai K, Ahaghotu C, Agurs-Collins T, Kittles RA: **Association between Serum 25-Hydroxy-Vitamin D and Aggressive Prostate Cancer in African American Men.** *Nutrients* 2016, **9**.
17. Murphy AB, Nyame Y, Martin IK, Catalona WJ, Hollowell CM, Nadler RB, Kozlowski JM, Perry KT, Kajdacsy-Balla A, Kittles R: **Vitamin D deficiency predicts prostate biopsy outcomes.** *Clin Cancer Res* 2014, **20**:2289-2299.
18. Hochberg Z, Templeton AR: **Evolutionary perspective in skin color, vitamin D and its receptor.** *Hormones (Athens)* 2010, **9**:307-311.
19. Jablonski NG: **The evolution of human skin colouration and its relevance to health in the modern world.** *J R Coll Physicians Edinb* 2012, **42**:58-63.
20. Yao S, Hong CC, Bandera EV, Zhu Q, Liu S, Cheng TD, Zirpoli G, Haddad SA, Lunetta KL, Ruiz-Narvaez EA, et al: **Demographic, lifestyle, and genetic determinants of circulating concentrations of 25-hydroxyvitamin D and vitamin D-binding protein in African American and European American women.** *Am J Clin Nutr* 2017, **105**:1362-1371.
21. Yao S, Haddad SA, Hu Q, Liu S, Lunetta KL, Ruiz-Narvaez EA, Hong CC, Zhu Q, Sucheston-Campbell L, Cheng TY, et al: **Genetic variations in vitamin D-related pathways and breast cancer risk in African American women in the AMBER consortium.** *Int J Cancer* 2016, **138**:2118-2126.
22. Mishra DK, Wu Y, Sarkissyan M, Sarkissyan S, Chen Z, Shang X, Ong M, Heber D, Koeffler HP, Vadgama JV: **Vitamin D receptor gene polymorphisms and prognosis of breast cancer among African-American and Hispanic women.** *PLoS One* 2013, **8**:e57967.
23. Neuhouwer ML, Bernstein L, Hollis BW, Xiao L, Ambis A, Baumgartner K, Baumgartner R, McTiernan A, Ballard-Barbash R: **Serum vitamin D and breast density in breast cancer survivors.** *Cancer Epidemiol Biomarkers Prev* 2010, **19**:412-417.
24. John EM, Schwartz GG, Koo J, Wang W, Ingles SA: **Sun exposure, vitamin D receptor gene polymorphisms, and breast cancer risk in a multiethnic population.** *Am J Epidemiol* 2007, **166**:1409-1419.
25. Layne TM, Weinstein SJ, Graubard BI, Ma X, Mayne ST, Albanes D: **Serum 25-hydroxyvitamin D, vitamin D binding protein, and prostate cancer risk in black men.** *Cancer* 2017.
26. Batai K, Murphy AB, Nonn L, Kittles RA: **Vitamin D and Immune Response: Implications for Prostate Cancer in African Americans.** *Front Immunol* 2016, **7**:53.
27. Richards Z, Batai K, Farhat R, Shah E, Makowski A, Gann PH, Kittles R, Nonn L: **Prostatic compensation of the vitamin D axis in African American men.** *JCI Insight* 2017, **2**:e91054.
28. Kidd LC, Paltoo DN, Wang S, Chen W, Akereyeni F, Isaacs W, Ahaghotu C, Kittles R: **Sequence variation within the 5' regulatory regions of the vitamin D binding protein and receptor genes and prostate cancer risk.** *Prostate* 2005, **64**:272-282.
29. Murphy AB, Kelley B, Nyame YA, Martin IK, Smith DJ, Castaneda L, Zagaja GJ, Hollowell CM, Kittles RA: **Predictors of serum vitamin D levels in African American and European American men in Chicago.** *Am J Mens Health* 2012, **6**:420-426.
30. Paller CJ, Kanaan YM, Beyene DA, Naab TJ, Copeland RL, Tsai HL, Kanarek NF, Hudson TS: **Risk of prostate cancer in African-American men: Evidence of mixed effects of dietary quercetin by serum vitamin D status.** *Prostate* 2015, **75**:1376-1383.
31. Taksler GB, Cutler DM, Giovannucci E, Smith MR, Keating NL: **Ultraviolet index and racial differences in prostate cancer incidence and mortality.** *Cancer* 2013, **119**:3195-3203.
32. Batai K, Kittles RA: **Can vitamin D supplementation reduce prostate cancer disparities?** *Pharmacogenomics* 2016, **17**:1117-1120.
33. King L, Dear K, Harrison SL, van der Mei I, Brodie AM, Kimlin MG, Lucas RM: **Investigating the patterns and determinants of seasonal variation in vitamin D status in Australian adults: the Seasonal D Cohort Study.** *BMC Public Health* 2016, **16**:892.

34. Diffey BL: **Modelling the seasonal variation of vitamin D due to sun exposure.** *Br J Dermatol* 2010, **162**:1342-1348.
35. Manson JE, Cook NR, Lee IM, Christen W, Bassuk SS, Mora S, Gibson H, Albert CM, Gordon D, Copeland T, et al: **Marine n-3 Fatty Acids and Prevention of Cardiovascular Disease and Cancer.** *N Engl J Med* 2019, **380**:23-32.
36. Bassuk SS, Chandler PD, Buring JE, Manson JE, Group VR: **The VITamin D and OmegA-3 Trial (VITAL): Do Results Differ by Sex or Race/Ethnicity?** *Am J Lifestyle Med* 2021, **15**:372-391.
37. Chandler PD, Chen WY, Ajala ON, Hazra A, Cook N, Bubes V, Lee IM, Giovannucci EL, Willett W, Buring JE, et al: **Effect of Vitamin D3 Supplements on Development of Advanced Cancer: A Secondary Analysis of the VITAL Randomized Clinical Trial.** *JAMA Netw Open* 2020, **3**:e2025850.
38. Marshall DT, Savage SJ, Garrett-Mayer E, Keane TE, Hollis BW, Horst RL, Ambrose LH, Kindy MS, Gattoni-Celli S: **Vitamin D3 supplementation at 4000 international units per day for one year results in a decrease of positive cores at repeat biopsy in subjects with low-risk prostate cancer under active surveillance.** *J Clin Endocrinol Metab* 2012, **97**:2315-2324.
39. Carlberg C, Campbell MJ: **Vitamin D receptor signaling mechanisms: integrated actions of a well-defined transcription factor.** *Steroids* 2013, **78**:127-136.
40. Kawai M, Kinoshita S, Yamazaki M, Yamamoto K, Rosen CJ, Shimba S, Ozono K, Michigami T: **Intestinal clock system regulates skeletal homeostasis.** *JCI Insight* 2019, **4**.
41. Teboul M, Guillaumond F, Grechez-Cassiau A, Delaunay F: **The nuclear hormone receptor family round the clock.** *Mol Endocrinol* 2008, **22**:2573-2582.
42. Trump DL, Potter DM, Muindi J, Brufsky A, Johnson CS: **Phase II trial of high-dose, intermittent calcitriol (1,25 dihydroxyvitamin D3) and dexamethasone in androgen-independent prostate cancer.** *Cancer* 2006, **106**:2136-2142.
43. Trump DL: **Calcitriol and cancer therapy: A missed opportunity.** *Bone Rep* 2018, **9**:110-119.
44. Long MD, Singh PK, Russell JR, Llimos G, Rosario S, Rizvi A, van den Berg PR, Kirk J, Sucheston-Campbell LE, Smiraglia DJ, Campbell MJ: **The miR-96 and RARgamma signaling axis governs androgen signaling and prostate cancer progression.** *Oncogene* 2019, **38**:421-444.
45. Ling MT, Chan KW, Choo CK: **Androgen induces differentiation of a human papillomavirus 16 E6/E7 immortalized prostate epithelial cell line.** *J Endocrinol* 2001, **170**:287-296.
46. Theodore S, Sharp S, Zhou J, Turner T, Li H, Miki J, Ji Y, Patel V, Yates C, Rhim JS: **Establishment and characterization of a pair of non-malignant and malignant tumor derived cell lines from an African American prostate cancer patient.** *Int J Oncol* 2010, **37**:1477-1482.
47. Hooker SE, Jr., Woods-Burnham L, Bathina M, Lloyd S, Gorjala P, Mitra R, Nonn L, Kimbro KS, Kittles RA: **Genetic Ancestry Analysis Reveals Misclassification of Commonly Used Cancer Cell Lines.** *Cancer Epidemiol Biomarkers Prev* 2019, **28**:1003-1009.
48. Papachristou EK, Kishore K, Holding AN, Harvey K, Roumeliotis TI, Chilamakuri CSR, Omarjee S, Chia KM, Swarbrick A, Lim E, et al: **A quantitative mass spectrometry-based approach to monitor the dynamics of endogenous chromatin-associated protein complexes.** *Nat Commun* 2018, **9**:2311.
49. Taberlay PC, Achinger-Kawecka J, Lun AT, Buske FA, Sabir K, Gould CM, Zotenko E, Bert SA, Giles KA, Bauer DC, et al: **Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations.** *Genome Res* 2016, **26**:719-731.
50. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK: **Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.** *Mol Cell* 2010, **38**:576-589.
51. Layer RM, Pedersen BS, DiSera T, Marth GT, Gertz J, Quinlan AR: **GIGGLE: a search engine for large-scale integrated genome analysis.** *Nat Methods* 2018, **15**:123-126.

52. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B: **A high-resolution map of the three-dimensional chromatin interactome in human cells.** *Nature* 2013, **503**:290-294.
53. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al: **Integrative analysis of 111 reference human epigenomes.** *Nature* 2015, **518**:317-330.
54. Jung I, Schmitt A, Diao Y, Lee AJ, Liu T, Yang D, Tan C, Eom J, Chan M, Chee S, et al: **A compendium of promoter-centered long-range chromatin interactions in the human genome.** *Nat Genet* 2019, **51**:1442-1449.
55. Qin Q, Fan J, Zheng R, Wan C, Mei S, Wu Q, Sun H, Brown M, Zhang J, Meyer CA, Liu XS: **Lisa: inferring transcriptional regulators through integrative modeling of public chromatin accessibility and ChIP-seq data.** *Genome Biol* 2020, **21**:32.
56. Breitkreutz BJ, Stark C, Tyers M: **The GRID: the General Repository for Interaction Datasets.** *Genome Biol* 2003, **4**:R23.
57. Wang S, Sun H, Ma J, Zang C, Wang C, Wang J, Tang Q, Meyer CA, Zhang Y, Liu XS: **Target analysis by integration of transcriptome and ChIP-seq data with BETA.** *Nat Protoc* 2013, **8**:2502-2515.
58. Siddappa M, Wani SA, Long MD, Leach DA, Mathe EA, Bevan CL, Campbell MJ: **Identification of transcription factor co-regulators that drive prostate cancer progression.** *Sci Rep* 2020, **10**:20332.
59. Niida H, Matsunuma R, Horiguchi R, Uchida C, Nakazawa Y, Motegi A, Nishimoto K, Sakai S, Ohhata T, Kitagawa K, et al: **Phosphorylated HBO1 at UV irradiated sites is essential for nucleotide excision repair.** *Nat Commun* 2017, **8**:16102.
60. Marshall JR, Tangen CM, Sakr WA, Wood DP, Jr., Berry DL, Klein EA, Lippman SM, Parnes HL, Alberts DS, Jarrard DF, et al: **Phase III trial of selenium to prevent prostate cancer in men with high-grade prostatic intraepithelial neoplasia: SWOG S9917.** *Cancer Prev Res (Phila)* 2011, **4**:1761-1769.
61. Batai K, Murphy AB, Ruden M, Newsome J, Shah E, Dixon MA, Jacobs ET, Hollowell CM, Ahaghotu C, Kittles RA: **Race and BMI modify associations of calcium and vitamin D intake with prostate cancer.** *BMC Cancer* 2017, **17**:64.
62. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, Schoech A, Pasaniuc B, Price AL: **Leveraging Polygenic Functional Enrichment to Improve GWAS Power.** *Am J Hum Genet* 2019, **104**:65-75.
63. Dorff T, Hirasawa Y, Acoba J, Pagano I, Tamura D, Pal S, Zhang M, Waitz R, Dhal A, Haynes W, et al: **Phase Ib study of patients with metastatic castrate-resistant prostate cancer treated with different sequencing regimens of atezolizumab and sipuleucel-T.** *J Immunother Cancer* 2021, **9**.
64. Fleet JC, Kovalenko PL, Li Y, Smolinski J, Spees C, Yu JG, Thomas-Ahner JM, Cui M, Neme A, Carlberg C, Clinton SK: **Vitamin D Signaling Suppresses Early Prostate Carcinogenesis in TgAPT121 Mice.** *Cancer Prev Res (Phila)* 2019, **12**:343-356.
65. Baumann B, Lugli G, Gao S, Zenner M, Nonn L: **High levels of PIWI-interacting RNAs are present in the small RNA landscape of prostate epithelium from vitamin D clinical trial specimens.** *Prostate* 2019, **79**:840-855.
66. Colston K, Colston MJ, Feldman D: **1,25-dihydroxyvitamin D3 and malignant melanoma: the presence of receptors and inhibition of cell growth in culture.** *Endocrinology* 1981, **108**:1083-1086.
67. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, Compton CC, DeLuca DS, Peter-Demchok J, Gelfand ET, et al: **A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project.** *Biopreserv Biobank* 2015, **13**:311-319.
68. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, et al: **Proteomics. Tissue-based map of the human proteome.** *Science* 2015, **347**:1260419.

69. Hahn J, Cook NR, Alexander EK, Friedman S, Walter J, Bubes V, Kotler G, Lee IM, Manson JE, Costenbader KH: **Vitamin D and marine omega 3 fatty acid supplementation and incident autoimmune disease: VITAL randomized controlled trial.** *BMJ* 2022, **376**:e066452.
70. Choi JH, Lee B, Lee JY, Kim CH, Park B, Kim DY, Kim HJ, Park DY: **Relationship between Sleep Duration, Sun Exposure, and Serum 25-Hydroxyvitamin D Status: A Cross-sectional Study.** *Sci Rep* 2020, **10**:4168.
71. Ramakrishnan S, Steck SE, Arab L, Zhang H, Bensen JT, Fontham ETH, Johnson CS, Mohler JL, Smith GJ, Su LJ, Woloszynska A: **Association among plasma 1,25(OH)₂ D, ratio of 1,25(OH)₂ D to 25(OH)D, and prostate cancer aggressiveness.** *Prostate* 2019, **79**:1117-1124.
72. Rosenberg A, Netter OS, Gogana P, Sheikh U, Macias V, Kajdacsy-Balla A, Sharifi R, Kittles RA, Murphy AB: **Physiologic serum 1,25 dihydroxyvitamin D is inversely associated with prostatic Ki67 staining in a diverse sample of radical prostatectomy patients.** *Cancer Causes Control* 2019, **30**:207-214.
73. Doig CL, Singh PK, Dhiman VK, Thorne JL, Battaglia S, Sobolewski M, Maguire O, O'Neill LP, Turner BM, McCabe CJ, et al: **Recruitment of NCOR1 to VDR target genes is enhanced in prostate cancer cells and associates with altered DNA methylation patterns.** *Carcinogenesis* 2013, **34**:248-256.
74. Thorne JL, Maguire O, Doig CL, Battaglia S, Fehr L, Sucheston LE, Heinaniemi M, O'Neill LP, McCabe CJ, Turner BM, et al: **Epigenetic control of a VDR-governed feed-forward loop that regulates p21(waf1/cip1) expression and function in non-malignant prostate cells.** *Nucleic Acids Res* 2011, **39**:2045-2056.
75. Battaglia S, Maguire O, Thorne JL, Hornung LB, Doig CL, Liu S, Sucheston LE, Bianchi A, Khanim FL, Gommersall LM, et al: **Elevated NCOR1 disrupts PPARalpha/gamma signaling in prostate cancer and forms a targetable epigenetic lesion.** *Carcinogenesis* 2010, **31**:1650-1660.
76. Khanim FL, Gommersall LM, Wood VH, Smith KL, Montalvo L, O'Neill LP, Xu Y, Peehl DM, Stewart PM, Turner BM, Campbell MJ: **Altered SMRT levels disrupt vitamin D3 receptor signalling in prostate cancer cells.** *Oncogene* 2004, **23**:6712-6725.
77. Solomon JD, Heitzer MD, Liu TT, Beumer JH, Parise RA, Normolle DP, Leach DA, Buchanan G, DeFranco DB: **VDR activity is differentially affected by Hic-5 in prostate cancer and stromal cells.** *Mol Cancer Res* 2014, **12**:1166-1180.
78. Ansems M, Karthaus N, Hontelez S, Aalders T, Looman MW, Verhaegh GW, Schalken JA, Adema GJ: **DC-SCRIPT: AR and VDR regulator lost upon transformation of prostate epithelial cells.** *Prostate* 2012, **72**:1708-1717.
79. Campbell FC, Xu H, El-Tanani M, Crowe P, Bingham V: **The yin and yang of vitamin D receptor (VDR) signaling in neoplastic progression: operational networks and tissue-specific growth control.** *Biochem Pharmacol* 2010, **79**:1-9.
80. Zhang Z, Kovalenko P, Cui M, Desmet M, Clinton SK, Fleet JC: **Constitutive activation of the mitogen-activated protein kinase pathway impairs vitamin D signaling in human prostate epithelial cells.** *J Cell Physiol* 2010, **224**:433-442.
81. Toropainen S, Vaisanen S, Heikkinen S, Carlberg C: **The down-regulation of the human MYC gene by the nuclear hormone 1alpha,25-dihydroxyvitamin D3 is associated with cycling of corepressors and histone deacetylases.** *J Mol Biol* 2010, **400**:284-294.
82. Rane JK, Erb HH, Nappo G, Mann VM, Simms MS, Collins AT, Visakorpi T, Maitland NJ: **Inhibition of the glucocorticoid receptor results in an enhanced miR-99a/100-mediated radiation response in stem-like cells from human prostate cancers.** *Oncotarget* 2016, **7**:51965-51980.
83. Gong F, Chiu LY, Cox B, Aymard F, Clouaire T, Leung JW, Cammarata M, Perez M, Agarwal P, Brodbelt JS, et al: **Screen identifies bromodomain protein ZMYND8 in chromatin recognition of transcription-associated DNA damage that promotes homologous recombination.** *Genes Dev* 2015, **29**:197-211.

84. Rafehi H, Orlowski C, Georgiadis GT, Ververis K, El-Osta A, Karagiannis TC: **Clonogenic assay: adherent cells.** *J Vis Exp* 2011.
85. Mohammed H, Taylor C, Brown GD, Papachristou EK, Carroll JS, D'Santos CS: **Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes.** *Nat Protoc* 2016, **11**:316-326.
86. Liao Y, Smyth GK, Shi W: **The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads.** *Nucleic Acids Res* 2019, **47**:e47.
87. Lun AT, Smyth GK: **csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows.** *Nucleic Acids Res* 2016, **44**:e45.
88. Ou J, Liu H, Yu J, Kelliher MA, Castilla LH, Lawson ND, Zhu LJ: **ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data.** *BMC Genomics* 2018, **19**:169.

Figure 1

Figure 2

A

B

C

Figure 3

A

B

Figure 4

A

B

Figure 5

A

B

C

D

E

Figure 6

A

B

C

Figure 7

bioRxiv preprint doi: <https://doi.org/10.1101/2022.01.31.478573>; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Supplementary Figure 1

Supplementary Figure 2

A

B

Supplementary Figure 3

Supplementary Figure 4

bioRxiv preprint doi: <https://doi.org/10.1101/2022.01.31.478573>; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Supplementary Figure 6

Supplementary Figure 7

Supplementary Figure 8

RC43N

A

B

Supplementary Figure 9

A

B

Supplementary Figure 10

A

B

C

TCGA PRAD cohort: 233 EA and 37 AA: Distinguish TMRPSS2 fusion

D

Rayford et al cohort: 556 EA and 596 AA

PMID: 34083737

Supplementary Figure 11

A

BAZ1A qPCR results

bioRxiv preprint doi: <https://doi.org/10.1101/2022.01.31.478573>; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

BAZ1A qPCR results

CC-control vector with ethanol treatment BC-BAZ1A overexpression vector with ethanol treatment
CD-control vector with Vitamin D treatment BD-BAZ1A overexpression vector with Vitamin D treatment

B

Supplementary Figure 12