Identification of genes required for enzalutamide resistance in castration-resistant prostate cancer cells *in vitro*

Sarah E. Kohrt^{1,2}, Wisam N. Awadallah^{2,3}, Robert A. Phillips III⁴, Thomas C. Case⁵, Renjie Jin⁵, Jagpreet S. Nanda^{2,3}, Xiuping Yu⁶, Peter E. Clark⁷, Yajun Yi⁸, Robert J. Matusik⁵, Philip D. Anderson⁴, and Magdalena M. Grabowska^{1,2,3,9,^}

¹ Department of Pharmacology, Case Western Reserve University, Cleveland, OH

² Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH

³ Department of Urology, Case Western Reserve University, Cleveland, OH

⁴ Biological Sciences, Salisbury University, Salisbury, MD

⁵ Department of Urology, Vanderbilt University Medical Center, Nashville, TN

⁶ Department of Biochemistry, Louisiana State University Health Shreveport, Shreveport, LA

⁷ Department of Urology, Levine Cancer Center/Atrium Health, Charlotte, NC

⁸ Quality, Safety & Risk Prevention, Vanderbilt University Medical Center, Nashville, TN

⁹ Department of Biochemistry, Case Western Reserve University, Cleveland, OH

Running title: Supporters of enzalutamide resistance in CRPC cells

Keywords: prostate cancer, castration-resistance, enzalutamide, resistance

Financial support: The authors would like to acknowledge funding from the Joe C. Davis Foundation (to RJM), the Vanderbilt Institute for Clinical and Translational Research (VICTR, to YY, PEC, and RJM). The Vanderbilt Institute for Clinical and Translational Research (VICTR) is funded by the National Center for Advancing Translational Sciences (NCATS) Clinical

Translational Science Award (CTSA) Program, Award Number 5UL1TR002243. The content of this manuscript solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We would also like to acknowledge the Case Research Institute, a joint venture between University Hospitals and Case Western Reserve University, start-up funds (to MMG), and the Cell and Molecular Biology Training Program (T32 GM 008056 to SEK).

Conflict of interest disclosure: The authors have nothing to disclose

Word count: 4,599

Total number of figures: 4 figures, 2 tables

^ Address all correspondence to:

Magdalena M. Grabowska
2123 Adelbert Road
Wood Research Tower; RTG00
Cleveland, OH 44106
Phone: 216-368-5736
Email: mmg126@case.edu

Abstract

Castration-resistant prostate cancer can be treated with the anti-androgen enzalutamide, but responses and duration of response are variable. To identify genes that support enzalutamide resistance, we performed a short hairpin RNA (shRNA) screen in the bone-homing, castration-resistant prostate cancer cell line, C4-2B. We identified eleven genes (*TFAP2C*, *CAD*, *SPDEF*, *EIF6*, *GABRG2*, *CDC37*, *PSMD12*, *COL5A2*, *AR*, *MAP3K11*, and *ACAT1*), whose loss resulted in decreased cell survival in response to enzalutamide. To validate our screen, we performed transient knockdowns in C4-2B and 22Rv1 cells and evaluated cell survival in response to enzalutamide. Through these studies, we validated three genes (*ACAT1*, *MAP3K11*, and *PSMD12*) as supporters of enzalutamide resistance *in vitro*. Although *ACAT1* expression is lower in metastatic castration-resistant prostate cancer samples versus primary prostate cancer samples, knockdown of *ACAT1* was sufficient to reduce cell survival in C4-2B and 22Rv1 cells. *MAP3K11* expression increases with Gleason grade, and the highest expression is observed in metastatic castration-resistant disease. Knockdown of *MAP3K11* reduced cell survival and pharmacologic inhibition of MAP3K11 with CEP-1347 in combination with enzalutamide resulted in a dramatic increase in cell death. This was associated with decreased phosphorylation of AR-Serine650, which is required for maximal AR activation. Finally, while *PSMD12* expression did not change during disease progression, knockdown of *PSMD12* resulted in decreased AR and AR splice variant expression, likely contributing to the C4-2B and 22Rv1 decrease in cell survival. Our study has therefore identified at least three new supporters of enzalutamide resistance in castration-resistant prostate cancer cells *in vitro*.

Introduction

Early stage prostate cancer is an androgen-dependent disease, and advanced prostate cancer is largely treated with androgen deprivation therapy, which targets androgen receptor (AR). Inevitably, tumors progress to castration-resistant prostate cancer. In castration-resistant

prostate cancer, AR signaling continues through multiple mechanisms including AR full length (AR-FL) over-expression, increases in androgen production, and induction of constitutively active AR splice variants (AR-V)[1]. Currently, abiraterone acetate and enzalutamide are standard therapies in metastatic castration-resistant prostate cancer[2, 3]. Enzalutamide is well tolerated and effective, extending castration-resistant prostate cancer patient survival by 5 months versus placebo after chemotherapy[4] and delaying initiation of chemotherapy by 18 months versus placebo[5]. Nevertheless, tumors progress to enzalutamide-resistant castration-resistant prostate cancer. How these tumors progress to resistance is an area of active investigation, but both AR-dependent and AR-independent mechanisms have been proposed.

The purpose of our studies was to identify genes that support resistance to enzalutamide and identify gene products (proteins) that once targeted could re-sensitize tumors to enzalutamide. We first selected a bone metastasis-derived castration-resistant prostate cancer model, C4-2B[6], which upregulates AR-FL and AR-V7 in response to continuous enzalutamide treatment[7]. We then performed a short hairpin RNA (shRNA) screen, comparing the response of vehicle-treated and enzalutamide-treated C4-2B cells and assessed which genes, when knocked down, resulted in increased cell death. Through this approach, we identified 11 genes (*TFAP2C*, *CAD*, *SPDEF*, *EIF6*, *GABRG2*, *CDC37*, *PSMD12*, *COL5A2*, *AR*, *MAP3K11*, and *ACAT1*), of which *ACAT1*, *MAP3K11*, and *PSMD12* appear the most promising in validation studies.

Methods:

shRNA screen: The 27k Module 1 DECIPHER lentiviral shRNA library (Collecta, DHPAC-M1-P) was transduced into C4-2B cells, targeting approximately 5,043 genes, whose gene products represent members of signal transduction pathways and approved drug targets. Cells were transduced with a pooled lentiviral shRNA library, generated per manufacturer's instructions, and then split into three groups. Cells in the first group were collected and represent the initial

population of shRNA quantity (3095 YY1-2). The cells from the second group served as an enzalutamide-negative control group (cell culture in the enzalutamide vehicle control [DMSO] for 6 days; YY3095 3-5), and cells from the third group were cultured in the presence of 100nM enzalutamide for 6 days (3095 YY6-8). Genomic DNA was isolated and shRNA was quantified by deep sequencing. The ratio of the abundance of each shRNA in the third group versus both control groups was calculated. A given shRNA was considered a “hit” if it showed at least a 2-fold abundance decrease relative to both enzalutamide-negative control and initial samples.

Computational workflow: We performed quality control using on the sequencing data using FastQC[8]. Using the DECIPHER Data and Pathway Analysis Software <http://www.decipherproject.net/software/> called Bar-code Analyzer and Deconvoluter, we converted raw high-throughput sequencing data into a summary file for subsequent processing. This file includes annotation for every identified gene in our screen. For most transcripts being silenced, there are five barcodes from two different shRNA molecules targeting the transcript. Next, we used EdgeR [9, 10] and Limma[11, 12] to generate differential expression based on gene (barcode) count. After verifying that all of our samples had a similar sequencing depth, we used the Limma-trend test to identify which barcodes are decreased in enzalutamide-treated versus vehicle treated and untreated cells. Additional information can be found in the bioinformatics supplement (Supplemental Methods).

Human clinical data: To assess the expression of our screen-identified genes during prostate cancer progression, we used the Prostate Cancer Transcription Atlas web tool (<http://www.thepcta.org/>), that includes 1,321 clinical specimens from 38 prostate cancer cohorts, detailed here[13]. Publication-ready images were downloaded. To evaluate the frequency of alterations of our putative drivers in metastatic castration-resistant prostate cancer, we used cBioPortal[14, 15] to interrogate a set of publically available data [16] with associated outcome

and gene expression data [17]. This data set also includes an AR activity score, which is a measure of how active AR is, as well as a neuroendocrine prostate cancer (NEPCa) score, a measure of how neuroendocrine-like the gene expression pattern is. For the i) gene expression comparison between naïve and exposed patients and ii) correlation with AR and NEPCa activity scores, raw data was downloaded and plotted using GraphPad Prism. All other plots were exported as is from cBioPortal.

siRNA transient transfections: 22Rv1 cells were purchased from American Type Culture Collection (ATCC, CRL-2505). C4-2B[6] cells were provided by Drs. Ruoxiang Wu and Leland Chung (Cedars-Sinai) and Short Tandem Repeat (STR) validated. Both cell lines were mycoplasma tested following expansion of stocks. C4-2B and 22Rv1 cells were cultured in RPMI 1640 with L-glutamine in 10% fetal bovine serum. Cells were transfected with two siRNAs (Ambion Silencer Select, Thermo Fisher, see Table 1) per gene in Lipofectamine RNAi Max transfection reagent (Invitrogen) and OptiMEM medium (Gibco, Life Technologies). For cell survival assays, cells were transfected with 1 pmol siRNA for 24 hours then cells were either treated with 1:1000 DMSO (vehicle) or 10µM enzalutamide for six days in RPMI medium, with a media change at day three. For protein expression studies, cells were transfected with 30 pmol siRNA for 3 days or 6 days.

Quantitative PCR (qPCR): Total RNA was extracted from cells using the RNeasy Plus Mini kit (Qiagen) according to manufacturer instructions. RNA was reverse transcribed to cDNA using Bio-Rad iScript cDNA Synthesis Kit following manufacturer instructions. Following reverse transcription, qPCR was performed to quantitate RNA levels in C4-2B and 22Rv1 cells and to establish efficacy of siRNA constructs. qPCR was performed using the SYBR™ GREEN PCR Master Mix kit (Bio-Rad). Relative expressions of total RNA were normalized to endogenous control Glyceraldehyde 3-phosphate dehydrogenase (GAPDH). All

samples were run with two technical replicates and six biological replicates. No reverse transcriptase controls and no transcript controls were run in duplicate along with each primer in every experiment. Unpublished primers were designed using Primer-BLAST[18] to span exon-exon junctions when possible and identify as many transcript isoforms as possible. All primers were tested for amplification efficiency by using a standard curve generated from four 1:10 serial dilutions of cDNA prior to experiments. Efficiency was calculated using the formula $E = -1 + 10^{(-1/\text{slope})}$ where E=Efficiency. All primers used reported >80% efficiency according to this method. Primers are listed in Table 2.

Cell survival assays: Cell survival in response to knockdowns and enzalutamide was assessed using crystal violet staining. The crystal violet staining protocol was used as previously described [19]. In brief, 1×10^4 cells were seeded per well in 100uL RPMI 1640 with L-glutamine in 10% fetal bovine serum and underwent described drug treatments with eight biological replicates. At the end of the treatment course, wells were fixed in 4% paraformaldehyde in PBS (Alfa Aesar) and stained with 0.05% crystal violet stain (Ricca). Cells were washed with deionized water and dried for 16-24 hours before being imaged. Destaining for quantification was performed using 10% acetic acid and absorbance was read at 590 nm using a spectrophotometer (SpectraMax iD3, Molecular Devices). An average value of non-targeting vehicle was generated, and each value was divided by this average value to generate a normalized fold-change cell survival reading.

Western blotting: Protein isolation was performed using working RIPA buffer (120mM NaCl, 50 mM pH 8.0 Tris, 0.5% NP-40, 1 mM EGTA) containing protease and phosphatase inhibitors (100 ug/mL PMSF, 1 mM NaOrVa, 50 ug/mL aprotinin, 50ug/mL leupeptin). Protein concentration was determined using a Bio-Rad protein concentration assay dye and absorbance read at 595 nm on a spectrophotometer (SpectraMax ID3, Molecular Devices). 20µg total protein in βmercaptoethanol-containing loading buffer was added per well onto 4-12% SDS-PAGE gels

(NuPAGE) and transferred to PVDF (Bio-Rad) or nitrocellulose (Amersham) membrane. Nitrocellulose membranes were stained for total protein using Ponceau S (Boston BioProducts). Membranes were blocked with 10% BSA or milk-fat in TBST for 1 hour at room temperature. Primary antibodies from Abcam for AR (ab74272), AR-V7 (ab198394), phosphorylated AR-Ser650 (ab47563), ACAT1 (ab168342), MAP3K11 (ab51068), PSDM12 (ab229930), and GAPDH (AM4300) were used at recommended concentrations in 2.5% milk-fat or 2.5% bovine serum albumin overnight at 4°C. Anti-rabbit and anti-mouse secondary antibodies (GE Healthcare) were used in 2.5% milk-fat for one hour at room temperature. Membranes were imaged via chemiluminescence reagents (Super Signal™ West Pico Plus, Thermo Fisher) on a digital imager (ChemiDoc™ Touch, Bio-Rad).

Drug treatments: Cell response to the MAP3K11 inhibitor CEP-1347 was assessed in crystal violet cell survival assays. 1×10^4 cells were seeded per well in 100uL RPMI 1640 with L-glutamine in 10% fetal bovine serum and allowed to settle. CEP-1347 (Tocris, Catalog no. 4924) reagent was resuspended in DMSO for 1mM stock. Working solutions of 80μM, 40μM, 20μM, and 10μM were made with DMSO, then diluted 1:100 in RPMI 1640 with L-glutamine in 10% fetal bovine serum for final concentrations of 800nM, 400nM, 200nM, and 100nM, respectively. Enzalutamide (MDV3100; Selleckchem CAS No. 915087-33-1) was resuspended in DMSO for a working concentration of 100mM then diluted 1:1000 in media. DMSO control media contained equal amounts of solvent as drug treatment samples. 100uL of drug-containing media was plated per well for six days with new drug treatments performed on the third day. Cell survival at the end of drug treatment course was assessed via crystal violet staining as described.

Statistical analysis: Statistical analysis was performed using GraphPad Prism. For analysis of cell survival experiments, following the removal of outliers using the ROUT method, vehicle to vehicle, enzalutamide to enzalutamide comparisons were made using the Kruskal-Wallis test as

data was not always normally distributed. Comparisons between gene expression of abiraterone and/or enzalutamide exposed and naïve patients used the non-parametric Kruskal-Wallis test as data was not normally distributed. Correlations between gene expression (FPKM capture) and AR and NEPCa activity scores were calculated using non-parametric Spearman correlations. Clinical data from the Prostate Cancer Transcription Atlas web tool included statistical analysis.

Results

To identify genes that support enzalutamide resistance in castration-resistant prostate cancer *in vitro*, we transduced C4-2B cells with a lentiviral shRNA library that targeted 5,043 genes whose products act in signal transduction and are drug targets. The products of these genes include kinases, enzymes, receptors, and other proteins that already have targeted therapies. C4-2B cells were transduced with the shRNA library and split into three groups, one collected at day zero, one treated with vehicle for six days, and one treated with enzalutamide for six days. We used high-throughput sequencing to quantitate the number of shRNA barcodes in each sample. We were predominantly interested in the identity of the shRNA barcodes that had decreased in abundance in the enzalutamide-treated cells versus non-treated and vehicle-treated cells, as the loss of these barcodes reflected the death of cells in which these genes were essential. Through this analysis we identified 11 genes whose silencing resulted in increased cell death in response to enzalutamide: *TFAP2C*, *CAD*, *SPDEF*, *EIF6* (ITGB4BP), *GABRG2*, *CDC37*, *PSMD12*, *COL5A2*, *AR*, *MAP3K11*, and *ACAT1* (Figure 1A). As prostate cancer is driven by AR, and enzalutamide targets AR, the presence of *AR* in our gene list was an encouraging sign. The fold change differences were more dramatic between shRNA counts from enzalutamide-treated cells versus the initial population of shRNAs (Figure 1B) as compared to shRNA counts from vehicle treated cells (Figure 1C). This likely reflects the importance of some of these genes in cell survival in general and/or in response to DMSO.

Our first step was to evaluate whether our 11 genes were expressed in castration-resistant prostate cancer. First, we used cBioPortal[14, 15] to interrogate a set of publically available data from metastatic castration-resistant prostate cancer patient samples[16] with outcome and gene expression data. Unsurprisingly, *AR* was the most frequently altered of these genes, with 59% of samples bearing amplifications and mutations (Figure 1D). For the other genes, alterations were rarer, occurring in 2.8-6% of cases, depending on the gene. With the exception of *ACAT1*, which has more deletions, most of our gene list alterations were amplifications.

We also examined whether our putative resistance supporter genes increased after abiraterone and enzalutamide treatment. For this analysis, we included patients who were exposed or naïve to abiraterone and enzalutamide, and excluded patients on treatment or whose exposure status was unknown from Abida *et al*[16]. Through this analysis we determined that *AR* expression was dramatically increased in patients who had been exposed to abiraterone and enzalutamide (Supplemental Figure 1), as reported [16]. However, there was no difference in gene expression of our putative resistance drivers in patients who had received abiraterone and enzalutamide versus patients who had not (Supplemental Figure 1). With the exception of *AR*, all of our genes exhibited gains and losses in both patient groups (Supplemental Figure 2, 3).

To validate our putative supporters of enzalutamide resistance *in vitro*, we turned to siRNA knockdown studies. To complement validation studies in C4-2B cells, we also used 22Rv1 cells, which express AR-V, including AR-V7 and AR-V9, and are enzalutamide resistant[20, 21]. We used two siRNA constructs per gene and evaluated how knockdown of each gene impacts enzalutamide sensitivity and full length AR (AR-FL) and AR-V7 levels. Through this analysis, we validated three candidate genes (*ACAT1*, *MAP3K11*, and *PSMD12*) in two cell lines as supporters of enzalutamide resistance *in vitro*.

In our validation studies, we first assessed whether AR regulated expression of our genes of interest (Supplemental Figure 4). Our two AR constructs both efficiently knocked down expression of AR, with construct #2 being more efficient (Figure 1 E, F and Supplemental Figure

4A). Importantly, both of our knockdown constructs targeted AR-V7. Although AR-V7 expression has been reported in C4-2B cells[22], under our conditions, the expression was minimal. We next evaluated the consequences of *AR* knockdown on our genes of interest. Knockdown of *AR* reduced gene expression of *MAP3K11*, but it did not reduce the expression of *ACAT1* or *PSMD12* (Supplemental Figure 4B, C, D). We also evaluated whether knockdown of *AR* altered protein expression of ACAT1, MAP3K11, and PSMD12, and observed no significant changes. Interestingly, on a gene expression level, knockdown of many of our genes resulted in dramatic changes in expression of other members of our gene list (Supplemental Figure 4).

Our first gene of interest, *ACAT1*, encodes Acetyl-CoA Acetyltransferase 1 (ACAT1). While gene expression for *ACAT1* does not increase between benign and primary prostate cancer, *ACAT1* expression decreases in metastatic castration-resistant prostate cancer versus primary prostate cancer (Figure 2A-C). While *ACAT1* message appears to decrease between primary and metastatic castration-resistant prostate cancer, on a protein level, ACAT1 expression reportedly increases[23, 24]. For *ACAT1*, both siACAT1 constructs were efficient, and knockdown of *ACAT1* did not dramatically alter AR or AR-V7 protein expression (Figure 2D, E). Over a six-day treatment time course, knockdown of *ACAT1* resulted in a dramatic reduction in cell survival in C4-2B cells (55-76% siACAT1 enzalutamide treated versus non-targeting enzalutamide treated cells) (Figure 2F). In 22Rv1 cells, this effect was more modest; achieving a 14% survival reduction versus non-targeting enzalutamide treated cells (Figure 2G). To evaluate how ACAT1 could be supporting enzalutamide resistance, we again turned to the Abida *et al* data set[16]. Using cBioPortal[14, 15], we selected metastatic castration-resistant prostate cancer patients who were naïve to abiraterone and enzalutamide treatment and compared their AR activity and neuroendocrine (NE/NEPCa) activity scores. In clinical samples, *ACAT1* expression correlates with the AR activity score (Figure 2H) but not the NE/NEPCa score (Figure 2I).

Our second gene of interest was *MAP3K11*, which encodes MAP3K11, also known as Mixed Lineage Kinase 3 (MLK3). *MAP3K11* expression increases dramatically during prostate

cancer progression (Figure 3A-C). Both of the MAP3K11 targeting constructs knocked down *MAP3K11* expression, with construct #1 being more efficient (Figure 3D, E). The more efficient *MAP3K11* knockdown resulted in more cell death in response to enzalutamide (Figure 3F, G). In C4-2B cells, knockdown of *MAP3K11* with construct #1 and treatment with enzalutamide resulted in a 73% decrease in cell survival versus non-targeting enzalutamide treated cells, while in 22Rv1 cells this reduction was more modest with a reduction of 28%. Consistent with previous reports, when we knocked down *MAP3K11*, we observed no decrease in AR[25] or AR-V7 expression, but we did see a loss of AR-Ser650 phosphorylation in 22Rv1 cells; this decrease was more pronounced with greater MAP3K11 ablation (Figure 3E). Phosphorylation of AR-Ser650 promotes maximal AR transactivation activity[26], suggesting MAP3K11 could be supporting enzalutamide resistance through activation of AR. Interestingly, in C4-2B cells, we did not observe AR-Ser650 phosphorylation under normal cell culture conditions, consistent with previous reports where AR-Ser650 phosphorylation is induced by phorbol 12-myristate 13-acetate (PMA)[25]. In the abiraterone and enzalutamide naïve patients, *MAP3K11* expression was not statistically correlated with the AR activity score (Figure 3H), but it was negatively correlated with the NE/NEPCA score (Figure 3I).

As several inhibitors have been developed for the MLK family, we next evaluated whether one of these could be used in combination with enzalutamide. CEP-1347 was initially developed to prevent HIV-1 Associated Neurocognitive Disorders (HAND) and Parkinson's[27-29]. Treatment of 22Rv1 cells with CEP-1347 did not reduce AR or AR-V7 expression, but it did reduce AR-Ser650 phosphorylation to similar levels as *MAP3K11* knockdown (Figure 3J). Indeed, combination therapy with increasing concentrations of CEP-1347 induced increased cell death in response to 10µM enzalutamide versus vehicle treated cells in C4-2B and 22Rv1 cells in a dose-dependent manner (Figure 3K, L).

Our final gene is *PSMD12*, which encodes Proteasome 26S Subunit, Non-ATPase 12, a component of the 26S Proteasome. *PSMD12* expression does not change between benign,

primary, and metastatic prostate cancer (Figure 4A-C). In our validation studies, both constructs were highly efficient at knocking down *PSMD12* gene expression in both cell lines (Figure 4D, E). In cell survival assays, we observed that C4-2B cells were exquisitely sensitive to *PSMD12* knockdown, with knockdown driving an 85-91% decrease in cell survival in C4-2B cells and a 28% reduction in survival in 22Rv1 cells (Figure 4F, G). This increased sensitivity was likely due to a loss of AR and AR-V7 expression, as when *PSMD12* expression decreased, AR and AR-V7 expression decreased (Figure 4C, D). In the clinical data *PSMD12* expression was not correlated with AR activity or NE/NEPCa activity scores (Figure 4E, F).

Discussion

Our shRNA screen in C4-2B cells identified 11 putative supporters of enzalutamide resistance (*TFAP2C*, *CAD*, *SPDEF*, *EIF6*, *GABRG2*, *CDC37*, *PSMD12*, *COL5A2*, *AR*, *MAP3K11*, and *ACAT1*). Our siRNA constructs efficiently targeted all of our genes of interest, and revealed that, at least on a gene expression level, there is considerable cross-talk between these genes. For example, knockdown of *MAP3K11*, results in decreased expression of genes that are regulated by AR, likely by reducing its transcriptional efficiency. Through our validation studies, we verified that transient knockdown of *ACAT1*, *MAP3K11*, or *PSMD12* could sensitize castration-resistant prostate cancer cells to enzalutamide.

In general, C4-2B cells were much more sensitive to knockdown of these three genes in combination with enzalutamide treatment. There are two possible explanations for this observation. The first is that we identified these genes as supporting enzalutamide resistance in C4-2B cells through our shRNA screen, and thus these cells are more dependent on these pathways. Alternatively, it is possible that the high level of AR-V in 22Rv1 cells, which confers enzalutamide resistance by removing dependency on ligand binding, simply makes 22Rv1 cells much more resistant. As knockdown of *PSMD12* reduces AR and AR-V7 expression, and 22Rv1 cells do not become dramatically more sensitive to enzalutamide, this would suggest some of

these genes might only be required for enzalutamide resistance in C4-2B cells, and by extension, only a subset of prostate cancers.

One of our enzalutamide-resistance supporting genes, *ACAT1*, encodes the protein ACAT1 that has several roles in the mitochondria. First, it serves as the final enzyme in isoleucine metabolism, converting 2-methylacetoacetyl-CoA into propionyl-CoA and acetyl-CoA[30]. ACAT1 also functions in ketone body metabolism in which its substrate is acetoacetyl-CoA[30]. Produced acetyl-CoA is shuttled into the Krebs cycle to be oxidized for energy production. More recently, ACAT1 has been implicated in regulating the pyruvate dehydrogenase complex, comprised of pyruvate dehydrogenase and PDH phosphatase, through acetylation[31]. Importantly, knockdown of ACAT1 activity is sufficient to push cancer cells from their preferred aerobic glycolysis, favored by cancer and proliferating cells, and into oxidative phosphorylation, the energy production mechanism favored by differentiated cells. Beyond its role in metabolism, ACAT1 dysregulation could alter acetyl-CoA levels. Notably, AR can be acetylated at AR-K618 [32] in the DNA binding domain and AR-K632, AR-K633[33] and AR-K630[34] in the hinge region, which supports increased AR transcriptional activity and increased tumor growth in xenograft models[32, 34]. Which of these mechanisms supports enzalutamide resistance remains to be determined.

In prostate, ACAT1 is associated with more aggressive prostate cancer and castration-resistant disease[23, 24]. These observations, combined with our experimental data, suggest that ACAT1 is expressed in castration-resistant prostate cancer and may support enzalutamide resistance, making it an attractive therapeutic target. ACAT1 activity can be inhibited by arecoline hydrobromide[31] or the FDA-approved drug sulfasalazine[35]. Unfortunately, neither arecoline hydrobromide nor sulfasalazine are specific to ACAT1.

Another gene of interest *MAP3K11*, encoding MAP3K11, perhaps holds the most therapeutic potential. MAP3K11 is a serine/threonine kinase[36], which via phosphorylation of MAP2Ks (MKKs), is a regulator of the mitogen-activated protein kinases (MAPKs), including JNK, ERK, and p38 (reviewed[37]). MAP3K11 can act both as a kinase and a scaffold for other

kinases[37]. Earlier studies identified MAP3K11 as a potent regulator of AR transcriptional activity in C4-2B cells through an RNAi phenotypic screen[25]. The link between MAP3K11 and AR activity appears to be driven by phosphorylation of serine 650, located in the hinge region of AR[33]. The hinge region is absent in most AR-Vs, including AR-V7, but it is present in a subset, including AR^{v567es} [33], suggesting this posttranslational modification could regulate the activities of some AR-Vs. Based on alanine-for-serine mutation studies using human AR in COS cells, phosphorylation of AR-Ser650 is required for optimal AR transactivation activity, as without AR-Ser650 phosphorylation, transcription of an AR-responsive mammary tumor promoter was reduced by 30%[26]. While it is possible that AR-Ser650 is a direct target of MAP3K11, *in vitro* kinase assays revealed AR-Ser650 is phosphorylated by JNK and p38 [38], which are downstream of MAP3K11. Which of these pathways are effectors of MAP3K11 in castration-resistant prostate cancer is currently under investigation.

Of our three genes, MAP3K11 has the most specific and tested therapeutics. Importantly, there are several inhibitors that are somewhat selective for MAP3K11 versus other mixed lineage kinase family members: CEP-1347 and URM-099. CEP-1347 was initially developed to prevent HAND and Parkinson's[27-29], and while safe and well-tolerated it failed to prevent Parkinson's progression in a Phase II clinical trial[39]. CEP-1347 is a fairly selective MAP3K11 inhibitor ($IC_{50} = 23$ nM[28]; $IC_{50} = 6$ nM[40]), with limited off-target effects on other MLK family members (MLK1 $IC_{50} = 38$ nM and MLK2 $IC_{50} = 51$ nM[28]; MLK1 $IC_{50} < 1$ nM and MLK2 $IC_{50} = 2$ nM[40]). URM-099 is a more specific MAP3K11 inhibitor ($IC_{50} = 14$ nM), with limited off-target effects on other MLK family members (MLK1 $IC_{50} = 19$ nM; MLK2 $IC_{50} = 42$ nM; and DLK $IC_{50} = 150$ nM)[40]. Our current studies are focusing on evaluating these compounds as a therapy in conjunction with enzalutamide in enzalutamide-resistant castration-resistant prostate cancer.

Our final gene, *PSMD12*, has not been previously implicated in prostate cancer, but dysregulation of the proteasome has been associated with prostate cancer. Inhibition of the

proteasome with therapeutics like bortezomib has been shown to increase the efficacy of first generation anti-androgens like bicalutamide by decreasing the expression of AR and AR-V [41] or induce sensitivity of AR-independent prostate cancer cells to etoposide[42]. Unfortunately, proteasome inhibitors, to date, have proven acutely toxic, and targeting PSMD12 with currently available therapeutics is unlikely to provide a favorable risk-to-benefit ratio.

Although not all of our candidate genes validated in our follow-up studies, it should be noted that many of these hold promise and may be better evaluated in other cell lines and assays. For example, *GABRG2* encodes the Gamma-Aminobutyric Acid Type A Receptor Gamma2 Subunit protein, which is a GABA receptor. Previous studies by Jin *et al* discovered *GABRG2* as a member of a prognostic 21-gene panel (NARP21) that predicted decreased overall cancer-specific survival and metastasis-free survival of patients with prostate cancer [43]. Similarly, *CAD*, which encodes carbamoyl-phosphate synthetase II, aspartate transcarbamylase, and dihydroorotase is associated with the synthesis of pyrimidine nucleotides, necessary for cell proliferation. *CAD* is regulated by the MAPK cascade[44], and indeed knockdown of *MAP3K11* reduces *CAD* gene expression. *CAD* also interacts with AR and fosters AR translocation into the nucleus, and is posited as an early marker of prostate tumor recurrence[45]. Finally *CDC37*, which directs kinases to the HSP90 complex, has been implicated in supporting prostate cancer cell growth via increasing AR activity and activation of kinases [46]. The failure to validate some of these potential supporters of enzalutamide resistance, therefore, is likely to be due to the limitations of our assay rather than their lack of importance in prostate cancer biology.

These studies have several limitations. First, we have focused on genes that are already expressed in castration-resistant prostate cancer cell lines rather than those that have emerged because of treatment with enzalutamide. We have also only focused on castration-resistant prostate cancer treated with enzalutamide, as at the initiation of our studies, enzalutamide treatment was not yet approved in the metastatic hormone-sensitive setting[47]. In addition, our study has been limited to 5,043 genes that are involved in signal transduction and are drug

targets, which has left out a considerable number of potential drivers, like transcription factors, that likely play an important role in enzalutamide resistance. We have also performed our validation experiments in the context of transient knockdown experiments *in vitro* and focused exclusively on AR-centric resistance mechanisms. It is likely that ACAT1, MAP3K11, and PSMD12 act beyond AR to support enzalutamide resistance. Subsequent studies will focus on defining these enzalutamide-resistance drivers more holistically, including both loss of function and gain of function experiments to delineate how these genes support enzalutamide resistance both *in vitro* and *in vivo*. In summary, our studies have identified 11 genes (*TFAP2C*, *CAD*, *SPDEF*, *EIF6*, *GABRG2*, *CDC37*, *PSMD12*, *COL5A2*, *AR*, *MAP3K11*, and *ACAT1*) that support enzalutamide resistance in castration-resistant C4-2B cells, and we have validated three of these genes in enzalutamide-resistance *in vitro* (*ACAT1*, *MAP3K11*, and *PSMD12*).

Acknowledgements: We would like to thank Jianghong Zhang, PhD, formerly of Vanderbilt University Medical Center, for technical assistance with the shRNA screen.

Table 1: siRNA constructs

Name	Cat No. #1	Cat No. #2
siNT	4390843	4390846
Name	Gene ID #1	Gene ID #2
siAR	s1539	s1538
siACAT1	s900	s901
siMAP3K11	s8816	s8814
siGABRG2	s532146	s532148
siPSMD12	s11417	s11418
siCDC37	s21967	s21965
siCOL5A2	s3310	s533841
siSPDEF	s195114	s24518
siCAD	s2139	s224811
siEIF6	s7586	s7587
siTFAP2C	s14011	s14009

Table 2: qPCR primers

Number	Name	Sequence (5'-3')	Source
1	AR-TOTAL Fwd	GCAGGCAAGAGCACTGAAGATA	[48]
2	AR-TOTAL Rev	CCTTTGGTGTAACTCCCTTGA	[48]
3	TFAP2C Fwd	CTCCACGACATGCCTCACC	
4	TFAP2C Rev	CATGGAAATGGGACCTTTGCG	
5	SPDEF Fwd	AACAAGGAGAAGGGCATCTTCA	
6	SPDEF Rev	TAATACTGGCGGATGGAGCG	
7	EIF6 Fwd	CTGGACACAACCAGCACAGA	
8	EIF6 Rev	GACTCAGGTGAGGCTGTCAAT	
9	CDC37 Fwd	ACCCACCGACGCAAAGTA	
10	CDC37 Rev	CATCCTTCTCATCGCCCGTC	
11	PSMD12 Fwd	CTGTTGATGAGTCCGAAGCCT	
12	PSMD12 Rev	TTGGATCCTTGGGTCTCTGGA	
13	COL5A2 Fwd	AACAGAGGGTCTCAGTTCGC	
14	COL5A2 Rev	TGCCCCCTTTGAGAACCACAG	
15	ACAT1 Fwd	AGGTGGGATGGAGAGCATGT	
16	ACAT1 Rev	AGCACAGCTGCCCATATGAAT	
17	GAPDH Fwd	TGCACCACCAACTGCTTAGC	[49]
18	GAPDH Rev	GGCATGGACTGTGGTCATGAG	[49]
19	CAD Fwd	CACGACACCTGAAAGACCCC	
20	CAD Rev	TTGGCTCCTCAGCTGGCAAA	
21	MAP3K11 Fwd	CCCCCAGCACTCAATGGTAA	
22	MAP3K11 Rev	CTGGATCAGCTTGGGCGTC	
23	GABRG2 Fwd	TGCTCTACACCCTAAGGTTGAC	
24	GABRG2 Rev	CAAGGGGCAGGAGTGTTTCAT	
25	AR-V7 Fwd	CCATCTTGTCTCTTCGGAAATGTTA	[50]
26	AR-V7 Rev	TTTGAATGAGGCAAGTCAGCCTTTCT	[50]

References:

- [1] A.N. Vis, F.H. Schröder, Key targets of hormonal treatment of prostate cancer. Part 1: the androgen receptor and steroidogenic pathways, *BJU Int.*, 104 (2009) 438-448.
- [2] H. Scher, T. Beer, C. Higano, A. Anand, M. Taplin, E. Efstathiou, D. Rathkopf, J. Shelkey, E. Yu, J. Alumkal, D. Hung, M. Hirmand, L. Seely, M. Morris, D. Danila, J. Humm, S. Larson, M. Fleisher, C. Sawyers, Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study, *Lancet*, 375 (2010) 1437-1446.
- [3] K. Fizazi, H.I. Scher, A. Molina, C.J. Logothetis, K.N. Chi, R.J. Jones, J.N. Staffurth, S. North, N.J. Vogelzang, F. Saad, P. Mainwaring, S. Harland, O.B. Goodman, C.N. Sternberg, J.H. Li, T. Kheoh, C.M. Haqq, J.S. de Bono, Abiraterone acetate for treatment of metastatic castration-resistant prostate cancer: final overall survival analysis of the COU-AA-301 randomised, double-blind, placebo-controlled phase 3 study, *The Lancet Oncology*, 13 (2012) 983-992.
- [4] H.I. Scher, K. Fizazi, F. Saad, M.-E. Taplin, C.N. Sternberg, K. Miller, R. de Wit, P. Mulders, K.N. Chi, N.D. Shore, A.J. Armstrong, T.W. Flaig, A. Fléchon, P. Mainwaring, M. Fleming, J.D. Hainsworth, M. Hirmand, B. Selby, L. Seely, J.S. de Bono, Increased Survival with Enzalutamide in Prostate Cancer after Chemotherapy, *N. Engl. J. Med.*, 367 (2012) 1187-1197.
- [5] T.M. Beer , A.J. Armstrong , D.E. Rathkopf , Y. Loriot , C.N. Sternberg , C.S. Higano , P. Iversen , S. Bhattacharya , J. Carles , S. Chowdhury , I.D. Davis , J.S. de Bono , C.P. Evans , K. Fizazi , A.M. Joshua , C.-S. Kim , G. Kimura , P. Mainwaring , H. Mansbach , K. Miller , S.B. Noonberg , F. Perabo , D. Phung , F. Saad , H.I. Scher , M.-E. Taplin , P.M. Venner , B. Tombal Enzalutamide in Metastatic Prostate Cancer before Chemotherapy, *N. Engl. J. Med.*, 371 (2014) 424-433.
- [6] G.N. Thalmann, P.E. Anezinis, S.-M. Chang, H.E. Zhau, E.E. Kim, V.L. Hopwood, S. Pathak, A.C. von Eschenbach, L.W.K. Chung, Androgen-independent Cancer Progression and Bone Metastasis in the LNCaP Model of Human Prostate Cancer, *Cancer Res.*, 54 (1994) 2577-2581.
- [7] C. Liu, W. Lou, Y. Zhu, N. Nadiminty, C.T. Schwartz, C.P. Evans, A.C. Gao, Niclosamide Inhibits Androgen Receptor Variants Expression and Overcomes Enzalutamide Resistance in Castration-Resistant Prostate Cancer, *Clin. Cancer Res.*, 20 (2014) 3198-3210.
- [8] S. Andrews, FastQC A Quality Control tool for High Throughput Sequence Data, (2010).
- [9] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics*, 26 (2010) 139-140.
- [10] A.T.L. Yunshun Chen, Davis J McCarthy, Xiaobei Zhou, Mark D Robinson, Gordon K Smyth, edgeR: Empirical Analysis of Digital Gene Expression Data in R, (2016) R package version 3.14.10.
- [11] S. G.K., limma: Linear Models for Microarray Data, in: C.V.J. Gentleman R., Huber W., Irizarry R.A., Dudoit S. (Ed.) *Statistics for Biology and Health*, Springer, New York NY2005.
- [12] G. Smyth, Y. Hu, M. Ritchie, J. Silver , J. Wettenhall, D. McCarthy, D. Wu, W. Shi, B. Phipson, A. Lun, N. Thorne, A. Oshlack, C. de Graaf, Y. Chen, M. Langaas, E. Ferkingstad, M. Davy, F. Pepin, D. Choi limma: Linear Models for Microarray Data, (2016) R package version 3.28.21.
- [13] S. You, B.S. Knudsen, N. Erho, M. Alshalalfa, M. Takhar, H. Al-deen Ashab, E. Davicioni, R.J. Karnes, E.A. Klein, R.B. Den, A.E. Ross, E.M. Schaeffer, I.P. Garraway, J. Kim, M.R. Freeman, Integrated classification of prostate cancer reveals a novel luminal subtype with poor outcome, *Cancer Res.*, (2016) canres.0902.2016.

- [14] J. Gao, B.A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S.O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, N. Schultz, Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal, *Sci. Signal.*, 6 (2013) pl1-pl1.
- [15] E. Cerami, J. Gao, U. Dogrusoz, B.E. Gross, S.O. Sumer, B.A. Aksoy, A. Jacobsen, C.J. Byrne, M.L. Heuer, E. Larsson, Y. Antipin, B. Reva, A.P. Goldberg, C. Sander, N. Schultz, The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data, *Cancer Discovery*, 2 (2012) 401-404.
- [16] W. Abida, J. Cyrta, G. Heller, D. Prandi, J. Armenia, I. Coleman, M. Cieslik, M. Benelli, D. Robinson, E.M. Van Allen, A. Sboner, T. Fedrizzi, J.M. Mosquera, B.D. Robinson, N. De Sarkar, L.P. Kunju, S. Tomlins, Y.M. Wu, D. Nava Rodrigues, M. Loda, A. Gopalan, V.E. Reuter, C.C. Pritchard, J. Mateo, D. Bianchini, S. Miranda, S. Carreira, P. Rescigno, J. Filipenko, J. Vinson, R.B. Montgomery, H. Beltran, E.I. Heath, H.I. Scher, P.W. Kantoff, M.-E. Taplin, N. Schultz, J.S. deBono, F. Demichelis, P.S. Nelson, M.A. Rubin, A.M. Chinnaiyan, C.L. Sawyers, Genomic correlates of clinical outcome in advanced prostate cancer, *Proceedings of the National Academy of Sciences*, (2019) 201902651.
- [17] Analysis of Metastatic Prostate Adenocarcinoma (SU2C/PCF Dream Team, PNAS 2019), Created October 15, 2019.
- [18] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, T.L. Madden, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, *BMC Bioinformatics*, 13 (2012) 134-134.
- [19] E.S. Dahl, R. Buj, K.E. Leon, J.M. Newell, Y. Imamura, B.G. Bitler, N.W. Snyder, K.M. Aird, Targeting IDH1 as a pro-senescent therapy in high-grade serous ovarian cancer, *Molecular Cancer Research*, (2019) molcanres.1233.2018.
- [20] J.L. Van Etten, M. Nyquist, Y. Li, R. Yang, Y. Ho, R. Johnson, O. Ondigi, D.F. Voytas, C. Henzler, S.M. Dehm, Targeting a Single Alternative Polyadenylation Site Coordinately Blocks Expression of Androgen Receptor mRNA Splice Variants in Prostate Cancer, *Cancer Res.*, 77 (2017) 5228-5235.
- [21] Y. Li, S.C. Chan, L.J. Brand, T.H. Hwang, K.A.T. Silverstein, S.M. Dehm, Androgen Receptor Splice Variants Mediate Enzalutamide Resistance in Castration-Resistant Prostate Cancer Cell Lines, *Cancer Res.*, 73 (2013) 483-489.
- [22] S. Haile, M.D. Sadar, Androgen receptor and its splice variants in prostate cancer, *Cellular and molecular life sciences : CMLS*, 68 (2011) 3971-3981.
- [23] P. Saraon, D. Trudel, K. Kron, A. Dmitromanolakis, J. Trachtenberg, B. Bapat, T. van der Kwast, K.A. Jarvi, E.P. Diamandis, Evaluation and prognostic significance of ACAT1 as a marker of prostate cancer progression, *The Prostate*, 74 (2014) 372-380.
- [24] P. Saraon, D. Cretu, N. Musrap, G.S. Karagiannis, I. Batruch, A.P. Drabovich, T. van der Kwast, A. Mizokami, C. Morrissey, K. Jarvi, E.P. Diamandis, Quantitative Proteomics Reveals That Enzymes of the Ketogenic Pathway Are Associated with Prostate Cancer Progression, *Mol. Cell. Proteomics*, 12 (2013) 1589-1601.
- [25] H. Whitworth, S. Bhadel, M. Ivey, M. Conaway, A. Spencer, R. Hernan, H. Holemon, D. Gioeli, Identification of kinases regulating prostate cancer cell growth using an RNAi phenotypic screen, *PLoS ONE*, 7 (2012) e38950-e38950.
- [26] E.M. Wilson, J.A. Kemppainen, Z.X. Zhou, Identification of three proline-directed phosphorylation sites in the human androgen receptor, *Mol. Endocrinol.*, 9 (1995) 605-615.

- [27] J.R. Mathiasen, B.A.W. McKenna, M.S. Saporito, G.D. Ghadge, R.P. Roos, B.P. Holskin, Z.-L. Wu, S.P. Trusko, T.C. Connors, A.C. Maroney, B.A. Thomas, J.C. Thomas, D. Bozyczko-Coyne, Inhibition of mixed lineage kinase 3 attenuates MPP⁺-induced neurotoxicity in SH-SY5Y cells, *Brain Res.*, 1003 (2004) 86-97.
- [28] A.C. Maroney, J.P. Finn, T.J. Connors, J.T. Durkin, T. Angeles, G. Gessner, Z. Xu, S.L. Meyer, M.J. Savage, L.A. Greene, R.W. Scott, J.L. Vaught, CEP-1347 (KT7515), a Semisynthetic Inhibitor of the Mixed Lineage Kinase Family, *J. Biol. Chem.*, 276 (2001) 25302-25308.
- [29] M.S. Saporito, E.M. Brown, M.S. Miller, S. Carswell, CEP-1347/KT-7515, an Inhibitor of c-jun N-Terminal Kinase Activation, Attenuates the 1-Methyl-4-Phenyl Tetrahydropyridine-Mediated Loss of Nigrostriatal Dopaminergic Neurons In Vivo, *J. Pharmacol. Exp. Ther.*, 288 (1999) 421-427.
- [30] S.H. Korman, Inborn errors of isoleucine degradation: A review, *Mol. Genet. Metab.*, 89 (2006) 289-299.
- [31] J. Fan, R. Lin, S. Xia, D. Chen, S.E. Elf, S. Liu, Y. Pan, H. Xu, Z. Qian, M. Wang, C. Shan, L. Zhou, Q.-Y. Lei, Y. Li, H. Mao, B.H. Lee, J. Sudderth, R.J. DeBerardinis, G. Zhang, T. Owonikoko, M. Gaddh, M.L. Arellano, H.J. Khoury, F.R. Khuri, S. Kang, P.W. Doetsch, S. Lonial, T.J. Boggon, W.J. Curran, J. Chen, Tetrameric Acetyl-CoA Acetyltransferase 1 Is Important for Tumor Growth, *Mol. Cell*, 64 (2016) 859-874.
- [32] J.S. DePaolo, Z. Wang, J. Guo, G. Zhang, C. Qian, H. Zhang, J. Zabaleta, W. Liu, Acetylation of androgen receptor by ARD1 promotes dissociation from HSP90 complex and prostate tumorigenesis, *Oncotarget*, 7 (2016) 71417-71428.
- [33] T. van der Steen, D.J. Tindall, H. Huang, Posttranslational modification of the androgen receptor in prostate cancer, *Int. J. Mol. Sci.*, 14 (2013) 14833-14859.
- [34] M. Fu, M. Rao, C. Wang, T. Sakamaki, J. Wang, D. Di Vizio, X. Zhang, C. Albanese, S. Balk, C. Chang, S. Fan, E. Rosen, J.J. Palvimo, O.A. Jänne, S. Muratoglu, M.L. Avantaggiati, R.G. Pestell, Acetylation of androgen receptor enhances coactivator binding and promotes prostate cancer cell growth, *Mol. Cell. Biol.*, 23 (2003) 8563-8575.
- [35] L.D. Faison, H.L. White, Sulfasalazine inhibits lyso-PAF: Acetyl-CoA acetyltransferase, *Prostaglandins*, 44 (1992) 245-249.
- [36] K.A. Gallo, M.R. Mark, D.T. Scadden, Z. Wang, Q. Gu, P.J. Godowski, Identification and characterization of SPRK, a novel src-homology 3 domain-containing proline-rich kinase with serine/threonine kinase activity, *J. Biol. Chem.*, 269 (1994) 15092-15100.
- [37] C. Rattanasinchai, K.A. Gallo, MLK3 Signaling in Cancer Invasion, *Cancers*, 8 (2016) 51.
- [38] D. Gioeli, B.E. Black, V. Gordon, A. Spencer, C.T. Kesler, S.T. Eblen, B.M. Paschal, M.J. Weber, Stress Kinase Signaling Regulates Androgen Receptor Phosphorylation, Transcription, and Localization, *Mol. Endocrinol.*, 20 (2006) 503-515.
- [39] Mixed lineage kinase inhibitor CEP-1347 fails to delay disability in early Parkinson disease, *Neurology*, 69 (2007) 1480-1490.
- [40] V.S. Goodfellow, C.J. Loweth, S.B. Ravula, T. Wiemann, T. Nguyen, Y. Xu, D.E. Todd, D. Sheppard, S. Pollack, O. Polesskaya, D.F. Marker, S. Dewhurst, H.A. Gelbard, Discovery, synthesis, and characterization of an orally bioavailable, brain penetrant inhibitor of mixed lineage kinase 3, *J. Med. Chem.*, 56 (2013) 8032-8048.
- [41] R. Jin, H. Yamashita, X. Yu, J. Wang, O.E. Franco, Y. Wang, S.W. Hayward, R.J. Matusik, Inhibition of NF-kappa B signaling restores responsiveness of castrate-resistant prostate cancer

cells to anti-androgen treatment by decreasing androgen receptor-variant expression, *Oncogene*, (2014).

[42] B. Aras, A. Yerlikaya, Bortezomib and etoposide combinations exert synergistic effects on the human prostate cancer cell line PC-3, *Oncology letters*, 11 (2016) 3179-3184.

[43] R. Jin, Y. Yi, F.E. Yull, T.S. Blackwell, P.E. Clark, T. Koyama, J.A. Smith, Jr., R.J. Matusik, NF- κ B gene signature predicts prostate cancer progression, *Cancer Res.*, 74 (2014) 2763-2772.

[44] L.M. Graves, H.I. Guy, P. Kozlowski, M. Huang, E. Lazarowski, R.M. Pope, M.A. Collins, E.N. Dahlstrand, H.S. Earp, 3rd, D.R. Evans, Regulation of carbamoyl phosphate synthetase by MAP kinase, *Nature*, 403 (2000) 328-332.

[45] A. Morin, L. Fritsch, J.R.R. Mathieu, C. Gilbert, B. Guarmit, V. Firlej, C. Gallou-Kabani, A. Vieillefond, N.B. Delongchamps, F. Cabon, Identification of CAD as an androgen receptor interactant and an early marker of prostate tumor recurrence, *The FASEB Journal*, 26 (2011) 460-467.

[46] P.J. Gray, M.A. Stevenson, S.K. Calderwood, Targeting Cdc37 Inhibits Multiple Signaling Pathways and Induces Growth Arrest in Prostate Cancer Cells, *Cancer Research*, 67 (2007) 11942-11950.

[47] A.J. Armstrong, R.Z. Szmulewitz, D.P. Petrylak, J. Holzbeierlein, A. Villers, A. Azad, A. Alcaraz, B. Alekseev, T. Iguchi, N.D. Shore, B. Rosbrook, J. Sugg, B. Baron, L. Chen, A. Stenzl, ARCHES: A Randomized, Phase III Study of Androgen Deprivation Therapy With Enzalutamide or Placebo in Men With Metastatic Hormone-Sensitive Prostate Cancer, *J. Clin. Oncol.*, 37 (2019) 2974-2986.

[48] M.M. Grabowska, S.M. Kelly, A.L. Reese, J.M. Cates, T.C. Case, J. Zhang, D.J. DeGraff, D.W. Strand, N.L. Miller, P.E. Clark, S.W. Hayward, R.M. Gronostajski, P.D. Anderson, R.J. Matusik, Nfib regulates transcriptional networks that control the development of prostatic hyperplasia, *Endocrinology*, 157 (2016) 1094-1109.

[49] M.M. Grabowska, A.D. Elliott, D.J. DeGraff, P.D. Anderson, G. Anumanthan, H. Yamashita, Q. Sun, D.B. Friedman, D.L. Hachey, X. Yu, J.H. Sheehan, J.-M. Ahn, G.V. Raj, D.W. Piston, R.M. Gronostajski, R.J. Matusik, NFI Transcription Factors Interact with FOXA1 to Regulate Prostate-Specific Gene Expression, *Mol. Endocrinol.*, 28 (2014) 949-964.

[50] E.S. Antonarakis, C. Lu, H. Wang, B. Lubber, M. Nakazawa, J.C. Roeser, Y. Chen, T.A. Mohammad, Y. Chen, H.L. Fedor, T.L. Lotan, Q. Zheng, A.M. De Marzo, J.T. Isaacs, W.B. Isaacs, R. Nadal, C.J. Paller, S.R. Denmeade, M.A. Carducci, M.A. Eisenberger, J. Luo, AR-V7 and Resistance to Enzalutamide and Abiraterone in Prostate Cancer, *N. Engl. J. Med.*, 371 (2014) 1028-1038.

Figure 1: Identification of genes that support enzalutamide resistance in C4-2B cells. **A.** Log-scaled shRNA barcode counts of differentially expressed shRNA barcodes (genes) in C4-2B cells in response to enzalutamide (EnzPos) and vehicle (EnzNeg). Cells collected at the start of the experiment (Initial) represent the starting pool of shRNA barcodes. ITGB4BP is now EIF6. **B/C.** Under-represented shRNA barcodes (genes) between enzalutamide-treated and untreated (**B**) and vehicle-treated (**C**) cells. Volcano plots of eleven genes that support enzalutamide resistance. **D.** Most supporters of enzalutamide resistance are amplified in metastatic castration-resistant prostate cancer. Oncoprint of 429 patients from Abdia et al[16] in cBioPortal[14, 15] evaluating frequency of amplifications, deletions, and mutations. **E/F.** AR does not strongly regulate ACAT1, MAP3K11, or PSMD12. Knockdown of AR for three or six days in C4-2B (**E**) and 22Rv1 (**F**) cells.

Figure 2: ACAT1 supports enzalutamide resistance in vitro. A/B/C. *ACAT1* expression decreases during progression to metastatic castration-resistant prostate cancer. Lollipop (A), box plot (B) and lineplot of mean trend (C) of *ACAT1* expression in benign prostate, prostate cancer, and metastatic castration-resistant prostate cancer patient samples. GS: Gleason score; mCRPC: metastatic castration-resistant prostate cancer. **D/E.** Knockdown of *ACAT1* does not strongly impact AR expression. C4-2B (D) and 22Rv1 (E) cells were transfected with one of two non-targeting siRNA (siNT) or *ACAT1*-targeting (siACAT1 #1 or #2) and *ACAT1*, AR, AR-V7 (in 22Rv1 cells) was evaluated with GAPDH used as a loading control. **F/G.** Knockdown of *ACAT1* in castration-resistant prostate cancer cells increases cell death in response to enzalutamide. C4-2B (F) and 22Rv1 (G) were transfected with one of two non-targeting siRNA (siNT) or *ACAT1*-targeting (siACAT1) siRNAs and challenged with 10μM DMSO (vehicle) or enzalutamide for six days. Comparisons between DMSO and enzalutamide using Kruskal-Wallis test with Dunn's multiple comparisons test. * P < 0.05; ** P < 0.01; *** P < 0.001; **** P < 0.0001. **H.** *ACAT1* expression positively correlates with AR activity metastatic castration-resistant prostate cancer patient tissues. Correlation between *ACAT1* mRNA expression and the AR activity score evaluated by Spearman correlation in 106 abiraterone and enzalutamide naïve metastatic castration-resistant prostate cancer patients. **I.** *ACAT1* expression does not correlate with NEPCa activity metastatic castration-resistant prostate cancer patient tissues. Correlation between *ACAT1* mRNA expression and the NEPCa activity score evaluated by Spearman correlation.

Figure 3: MAP3K11 supports enzalutamide resistance in metastatic castration-resistant prostate cancer. Lollipop (A), box plot (B) and lineplot of mean trend (C) of *MAP3K11* expression in benign prostate, prostate cancer, and metastatic castration-resistant prostate cancer patient samples. GS: Gleason score; mCRPC: metastatic castration-resistant prostate cancer. **D/E.** Knockdown of *MAP3K11* does not strongly impact AR expression but it does reduce AR-Serine 650 (AR-Ser650) phosphorylation. C4-2B (D) and 22Rv1 (E) cells were transfected with one of two non-targeting siRNA (siNT) or *MAP3K11*-targeting (siMAP3K11 #1 or #2) and ACAT1, AR, AR-V7 (in 22Rv1 cells), and phosphorylated AR-Ser650 was evaluated with GAPDH used as a loading control. **F/G.** Knockdown of *MAP3K11* in castration-resistant prostate cancer cells increases cell death in response to enzalutamide. C4-2B (F) and 22Rv1 (G) were transfected with one of two non-targeting siRNA (siNT) or *MAP3K11*-targeting (siMAP3K11) siRNAs and challenged with 10 μ M DMSO (vehicle) or enzalutamide for six days. Comparisons between DMSO and DMSO, enzalutamide to enzalutamide using Kruskal-Wallis test with Dunn's multiple comparisons test. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$. **H.** *MAP3K11* expression does not correlate with AR activity metastatic castration-resistant prostate cancer patient tissues. Correlation between *MAP3K11* mRNA expression and the AR activity score evaluated by Spearman correlation in 106 abiraterone and enzalutamide naïve metastatic castration-resistant prostate cancer patients. **I.** *MAP3K11* expression inversely correlates with NEPCa activity metastatic castration-resistant prostate cancer patient tissues. Correlation between *MAP3K11* mRNA expression and the NEPCa activity score evaluated by Spearman correlation. **J.** Knockdown or inhibition of *MAP3K11* reduces AR-Ser650 phosphorylation. Comparison of AR-Ser650 phosphorylation in 22Rv1 cells transfected with non-targeting siRNA (siNT) or *MAP3K11* targeting siRNA (siMAP3K11) to vector or *MAP3K11* inhibitor (CEP-1347) treated cells. Importantly, AR and AR-V levels do not change. **K/L.** Inhibition of *MAP3K11* with CEP-1347 potentiates enzalutamide treatment. Treatment of C4-2B (K) and 22Rv1 (L) with increasing concentrations of CEP-1347 with 10 μ M DMSO (vehicle) or enzalutamide. Analysis by one-way ANOVA with Sidak's multiple comparisons test comparing 0nM CEP-1347 (DMSO) to varying CEP-1347 concentrations (black to black) and 0nM CEP-1347 (DMSO) plus enzalutamide to varying CEP-1347 concentrations plus enzalutamide (red to red). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$

Figure 4: PSMD12 supports enzalutamide resistance in vitro. **A/B/C.** *PSMD12* expression is similar in benign, primary, and metastatic castration-resistant prostate cancer. Lollipop (**A**), box plot (**B**) and lineplot of mean trend (**C**) of *PSMD12* expression in benign prostate, prostate cancer, and metastatic castration-resistant prostate cancer patient samples. GS: Gleason score; mCRPC: metastatic castration-resistant prostate cancer. **D/E.** Knockdown of *PSMD12* decreases AR and AR-V expression. C4-2B (**D**) and 22Rv1 (**E**) cells were transfected with one of two non-targeting siRNA (siNT) or *PSMD12*-targeting (siPSMD12 #1 or #2) and *PSMD12*, AR, AR-V7 (in 22Rv1 cells) was evaluated with GAPDH used as a loading control. **F/G.** Knockdown of *PSMD12* in castration-resistant prostate cancer cells increases cell death in response to enzalutamide. C4-2B (**F**) and 22Rv1 (**G**) were transfected with one of two non-targeting siRNA (siNT) or *PSMD12*-targeting (siPSMD12) siRNAs and challenged with 10 μ M DMSO (vehicle) or enzalutamide for six days. Comparisons between DMSO and DMSO, enzalutamide to enzalutamide using Kruskal-Wallis test with Dunn's multiple comparisons test. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$. **H/I.** *PSMD12* expression does not correlate with AR activity or NEPCa activity in 106 abiraterone and enzalutamide naïve metastatic castration-resistant prostate cancer patient tissues. Correlation between *PSMD12* mRNA expression and the AR activity score (**H**) or NEPCa activity (**I**) evaluated by Spearman correlation.

Supplemental Figure 1: Most putative enzalutamide-resistance drivers are not increased in response to abiraterone and enzalutamide treatment in patient samples. Gene expression changes of our putative drivers in metastatic castration-resistant prostate cancer in response to abiraterone and enzalutamide. Analysis of publically available data [16] with associated outcome and gene expression data [17], using cBioPortal[14, 15] interrogated using the Mann-Whitney U test. **** P < 0.0001.

Supplemental Figure 2: Genetic alterations in putative resistance drivers based on abiraterone and enzalutamide status. Gene expression changes overlapped with gene mutations/amplifications/deletions of our putative drivers in metastatic castration-resistant prostate cancer in response to abiraterone and enzalutamide. Analysis of publicly available data [16] with associated outcome and gene expression data [17], using cBioPortal[14, 15]. VUS: variants of unknown significance

AR: Putative copy-number alterations

Abiraterone (ABI) and Enzalutamide (ENZA) Exposure Status

ACAT1: Putative copy-number alterations

Abiraterone (ABI) and Enzalutamide (ENZA) Exposure Status

MMP3K11: Putative copy-number alterations

Abiraterone (ABI) and Enzalutamide (ENZA) Exposure Status

PSMD2: Putative copy-number alterations

Abiraterone (ABI) and Enzalutamide (ENZA) Exposure Status

TFAP2C: Putative copy-number alterations

Abiraterone (ABI) and Enzalutamide (ENZA) Exposure Status

CAD: Putative copy-number alterations

Abiraterone (ABI) and Enzalutamide (ENZA) Exposure Status

SPDEF: Putative copy-number alterations

Abiraterone (ABI) and Enzalutamide (ENZA) Exposure Status

EEF6: Putative copy-number alterations

Abiraterone (ABI) and Enzalutamide (ENZA) Exposure Status

GABRG2: Putative copy-number alterations

Abiraterone (ABI) and Enzalutamide (ENZA) Exposure Status

COL5A2: Putative copy-number alterations

Abiraterone (ABI) and Enzalutamide (ENZA) Exposure Status

CDCC3: Putative copy-number alterations

Abiraterone (ABI) and Enzalutamide (ENZA) Exposure Status

Supplemental Figure 3: Breakdown of genetic alterations by category in putative resistance drivers based on abiraterone and enzalutamide status. Gene alterations of our putative drivers in metastatic castration-resistant prostate cancer in response to abiraterone and enzalutamide. Analysis of publicly available data [16] with associated outcome and gene expression data [17] using cBioPortal[14, 15].

Supplemental Figure 4: Verification of knockdown of enzalutamide-resistance drivers. Knockdown of our eleven target genes (A-L) in C4-2B (i) and 22Rv1 (ii) cells. C4-2B cells do not express *GABRG2* or *AR-V7* at sufficient levels for detection. Data presented at log fold change.

Identification of Genes Promoting Resistance to Enzalutamide

Supplemental Methods To:

"Identification of genes required for enzalutamide resistance in castration-resistant prostate cancer cells in vitro"

Sarah E. Kohrt^{1,2}, Wisam N. Awadallah^{2,3}, Robert A. Phillips III⁴, Tom C. Case⁵, Renjie Jin⁵, Jagpreet S. Nanda^{2,3}, Xiuping Yu⁶, Peter E. Clark⁷, Yajun Yi⁸, Robert J. Matusik⁵, Philip D. Anderson⁴, and Magdalena M. Grabowska^{1,2,3,9}

¹ Department of Pharmacology, Case Western Reserve University, OH; ² Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH; ³ Department of Urology, Case Western Reserve University, Cleveland, OH; ⁴ Biological Sciences, Salisbury University, MD; ⁵ Department of Urology, Vanderbilt University Medical Center, Nashville, TN; ⁶ Department of Biochemistry, Louisiana State University Health Shreveport, Shreveport, LA; ⁷ Department of Urology, Carolinas HealthCare System, Charlotte, NC; ⁸ Quality, Safety & Risk Prevention, Vanderbilt University Medical Center, TN; ⁹ Department of Biochemistry, Case Western Reserve University, Cleveland, OH

Prepared by Robert A. Phillips III and Philip D. Anderson

February 7, 2017

Introduction

In 2012, the US Food and Drug Administration approved enzalutamide (Xtandi) for the treatment of patients with metastatic castration-resistant prostate cancer who have previously received docetaxel. In most patients, enzalutamide halts progression of the disease for a short period of time, but patients usually relapse, which is associated with poor prognosis. Identifying the genes and pathways that the cancer cells use to overcome enzalutamide are of interest, because new interventions aimed at preventing relapse after enzalutamide therapy would have therapeutic value to late-stage prostate cancer patients.

In this study, we took advantage of RNA interference techniques and an *in vitro* model of castration-resistant prostate cancer cells to identify genes that drive resistance to enzalutamide. We transduced one-third of a commercially available bar-coded lentiviral shRNA library (DECIPHER from Collecta) into C4-2B cells. The shRNA library contained shRNA clones targeting approximately 5045 human genes. The shRNA integrated into the chromosomal DNA of the transduced C4-2B cells and stably knocked down messenger RNA transcripts.

Our methodology called for the establishment of three populations of transduced C4-2B cells, which were maintained for six days before the cells were harvested and genomic DNA extracted. The first population (hereafter, "Initial") is a control that was maintained in the absence of enzalutamide and DMSO, which is the solvent for enzalutamide. The second population ("EnzNeg") was maintained in the presence of DMSO, but not enzalutamide, to control for the effects of the solvent. The third population ("EnzPos") was a test population maintained in the presence of both DMSO and enzalutamide.

Genomic DNA from 2 biological replicates from Initial and 3 biological replicates from each of EnzNeg and EnzPos were subjected to targeted deep sequencing on an Illumina instrument (total of eight samples). The targeting protocol promoted specifically the sequencing of the bar-coded shRNA in the genome.

In this analysis, we count the number of times each shRNA bar-code was sequenced in the eight samples, and calculate the ratio of the abundance of each shRNA in EnzPos versus both control groups. Our hypothesis is that if a shRNA-targeted gene helps cancer cells to achieve resistance when selected by enzalutamide, we will observe fewer bar-codes for those genes relative to both controls (which translates into more mRNA for the gene in EnzPos than both controls). A given shRNA was considered a "hit" if it showed at least two-fold abundance decrease in enzalutamide-positive relative to both enzalutamide-negative control and initial samples.

1 Report Linux Shell Variables

This first portion of the shRNA data analysis is coded in Linux bourne-again shell using freely available command-line tools. We declared the following directories as shell variables to condense and simplify the code for this analysis. The variables' paths were set in knitr's setup chunk using `Sys.setenv()`:

```
echo $Proj_Dir
echo $Raw_Dir
echo $Data_Dir
echo $Qual_Dir
echo $Sample_Dir
echo $Counts_Dir

## /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA
## /opt/C42B_shRNA/raw_data
## /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data
## /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/QC
## /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/Samples
## /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/Counts
```

2 Sample Preparation

Here we extract the compressed raw data using the `unzip` command from the Ubuntu Linux *utils* package. The code tests if the uncompressed data directory already exists and if so, deletes it, in keeping with the spirit of reproducible research. The uncompressed data directory is recreated by the `unzip` command without the need to invoke `mkdir`. The `-j` argument directs `unzip` to junk the archive's directory structure and dump all the files into the data (`-d`) directory.

```
if [ -d "$Data_Dir" ]; then
    rm -rf $Data_Dir
fi
unzip -j $Raw_Dir/C42B_shRNA.zip -d $Data_Dir

## Archive: /opt/C42B_shRNA/raw_data/C42B_shRNA.zip
## inflating: /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/2015-01-16_gsr15YY3095-1.p
## inflating: /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/3095-YY-1_1_sequence.txt.gz
## inflating: /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/3095-YY-2_1_sequence.txt.gz
## inflating: /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/3095-YY-3_1_sequence.txt.gz
## inflating: /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/3095-YY-4_1_sequence.txt.gz
## inflating: /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/3095-YY-5_1_sequence.txt.gz
## inflating: /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/3095-YY-6_1_sequence.txt.gz
## inflating: /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/3095-YY-7_1_sequence.txt.gz
## inflating: /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/3095-YY-8_1_sequence.txt.gz
## inflating: /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/Qubit DNA BR assay_3095-YY-
## inflating: /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/shRNA screen details.docx
## inflating: /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/Comments.xps
```

2.1 Quality Control

Quality control on the sequences is performed using FastQC [1]. This code runs FastQC for each of the eight sample files, and saves the results to the QC directory for us to inspect. By default, FastQC prints progress to the console, which is suppressed here with the `-quiet` argument. The `-threads 8` argument

directs FastQC to run in parallel for each of the eight samples, one CPU core each, without needing to invoke GNU parallel. Setting `--threads` to a number greater than eight does not make the code run faster. Linux `mkdir` is necessary because FastQC will not create the output directory (`--outdir`) if it does not already exist.

```
mkdir --mode 770 $Qual_Dir
fastqc --version
fastqc --quiet --outdir $Qual_Dir --threads 8 $Data_Dir/*.gz

## FastQC v0.11.2
```

FastQC showed that the first eighteen nucleotides of the 50-nucleotide sequences were high quality. These eighteen nucleotides are the CELLECTA barcodes that are the subject of this analysis.

3 Analysis

To analyze the barcodes, we will employ an in-house script that retains the first eighteen nucleotides from each sequence.

```
mkdir --mode 700 $Sample_Dir
cd $Sample_Dir
awk --version | head -n 1
parallel --version | head -n 1
parallel --jobs 8 'zcat {} | awk 'NR%4==2' | cut -c -18 > ${basename $(echo {} | sed
s/.txt.gz/.dna/)}' ::: $Data_Dir/*.gz

## GNU Awk 4.1.3, API: 1.1 (GNU MPFR 3.1.4, GNU MP 6.1.0)
## GNU parallel 20150522
```

The shRNA barcodes were tallied using the `table()` function in R. The following code was run in an Rscript to facilitate running using GNU parallel. This code is the content of the Rscript:

```
#!/usr/bin/env Rscript
args<-commandArgs(trailingOnly=T)
f<-file("stdin")
open(f)
MyInput<-as.character(readLines(f))
Tab<-as.data.frame(table(MyInput))
save(Tab,file="/tmp/Tab.Robj")
KV<-"/home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/DecipherKeyValuePairs.tsv"
Key<-read.table(file=KV,header=T,sep="\t")
tmp<-merge(x=Key,y=Tab,by.x="Barcode.Sequence",by.y="MyInput",all.x=T,all.y=F,sort=F)
tmp<-tmp[order(with(tmp,Col2)),] # Crucial to solve out-of-order problems caused by NA
write.table(x=tmp,
            file=paste("/home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/Counts/",
                      "S",args[1],sep=""),
            quote=F,append=F,sep="\t")
```

Linux inherits the Unix trait that disallows use of unexported variables in shells started from the running shell (as in `bash -c`). GNU parallel's `env_parallel` argument is still in beta testing. Therefore we use the full paths in the command below, which renders it a bit ugly. Running eight iterations of `Count_shRNA.R` on eight CPU cores (one core for each sample). This code takes about 45 seconds to run:


```
mkdir $Counts_Dir
parallel --jobs 8 'cat /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/Samples/3095-YY-{}_1
| Rscript --vanilla /home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Count_shRNA.R
{}' ::: {1..8}
```

Below is an example of the results from sample 1 for the Vimentin mRNA transcript. Vimentin chosen because the gene has a short name and most of the data will fit on this page.

```
sed -n 26934,26938p $Counts_Dir/S1
## 26935 GTACCAACGTACTGGTCA VIM|Homo sapiens vimentin (VIM), mRNA.|FLJ36605|7431|NM_003380.2|62414288|NP_00337
137
## 26933 ACACCAACGTACTGGTAC VIM|Homo sapiens vimentin (VIM), mRNA.|FLJ36605|7431|NM_003380.2|62414288|NP_00337
1747
## 26936 CAACCAACGTACTGGTGT VIM|Homo sapiens vimentin (VIM), mRNA.|FLJ36605|7431|NM_003380.2|62414288|NP_00337
2453
## 26934 TGACCAACGTACTGGTTG VIM|Homo sapiens vimentin (VIM), mRNA.|FLJ36605|7431|NM_003380.2|62414288|NP_00337
586
## 26937 ACTGCAACGTACTGGTTG VIM|Homo sapiens vimentin (VIM), mRNA.|FLJ36605|7431|NM_003380.2|62414288|NP_00337
1012
```

The data is arranged in four columns. The first column is an arbitrary row number that can be ignored. The second column contains the shRNA barcode against the transcript being silenced. The third column is annotation data that informs the transcript identity, NCBI database identifiers and cytoband information. The fourth column indicates the frequency at which the barcode was detected by deep sequencing the library prepared from this sample. For most transcripts being silenced, there are five barcodes from two different shRNA molecules targeting the transcript. We plan to compare barcode frequency between different samples to test the effect of enzalutamide on silencing. The next step is to combine all the counts data into a single file:

```
cd $Counts_Dir
paste S1 S2 S3 S4 S5 S6 S7 S8 | cut -f 1-4,8,12,16,20,24,28,32 > TotalCounts.tsv
```

4 Analysis of shRNA counts in R

The remained of the workflow is coded in the R language for statistical analysis. Here we import the data into R, ignoring the header line:

```
Path<-"/home/andersp/Research/RProj_Collab_with_Magdalena/C42B_shRNA/Data/Counts/TotalCounts.tsv"
Shrna<-read.table(file=Path,
                  sep="\t",
                  header=F,
                  stringsAsFactors=F,
                  skip=1,
                  comment.char="")
colnames(Shrna)<-c("Row", "Barcode", "Annotation",
                  "Initial.1", "Initial.2", "EnzN.1", "EnzN.2", "EnzN.3",
                  "EnzP.1", "EnzP.2", "EnzP.3")
```

At least one sample contains NA values which are actually zeroes. A zero means that the barcode in question was never found in the raw sequencing data. Here we coerce the NAs to zeroes:

```
Shrna[is.na(Shrna)]<-0
```

Further processing:

```
require(tidyr)

## Loading required package: tidyr

require(plyr)

## Loading required package: plyr

Shrna<-separate(data=Shrna,
               col=Annotation,
               into=c("Symbol", "Name", "Alt.Names", "EntrezID", "GenbankID",
                     "NoIdea", "NCBI.Protein.ID", "Chromosome", "Cytoband"),
               sep="\\|",
               extra="drop")

Shrna1<-ddply(.data=Shrna,
             .(Symbol),
             .drop=F,
             .fun=colwise(sum, .cols=c("Initial.1", "Initial.2", "EnzN.1", "EnzN.2",
                                       "EnzN.3", "EnzP.1", "EnzP.2", "EnzP.3")))
```

5 Differential Counts Analysis

EdgeR [2] and Limma [3] are R-Bioconductor packages for differential expression based on count data, which can be adapted for our purposes. Here we prepare the samples for analysis by creating a convenient object to store the count information and sample identities, called a DGEList.

```
require(edgeR)

## Loading required package: edgeR
## Loading required package: limma

rownames(Shrna1)<-Shrna1[["Symbol"]]
Shrna_List<-DGEList(counts=subset(Shrna1, select=-c(Symbol)),
                  group=c(rep("Initial", times=2),
                          rep("EnzNeg", times=3),
                          rep("EnzPos", times=3))
                  )
)
```

```
require(xtable)

## Loading required package: xtable

xtable(with(Shrna_List, samples), digits=0)
```

Next we apply a filter to remove genes from the analysis if they consistently have zero or very few counts. The method used in the edgeR vignette is to keep only those genes that have at least one read per million in at least 3 samples, because such low counts would make the differential counts analysis troublesome.

The following command determines the number of counts per million reads sequenced in each sample

	group	lib.size	norm.factors
Initial.1	Initial	22141545	1
Initial.2	Initial	19125653	1
EnzN.1	EnzNeg	20202085	1
EnzN.2	EnzNeg	19523160	1
EnzN.3	EnzNeg	20178264	1
EnzP.1	EnzPos	21474089	1
EnzP.2	EnzPos	20831135	1
EnzP.3	EnzPos	22619107	1

(cpm) and tests if the number is greater than one. If it is, then a TRUE result will be produced. The rowSums command is a convenience that has the same functionality as apply by rows: it coerces the TRUE/FALSE output to one or zero, then adds all the numbers together. The maximum number that can be produced for each gene in this study is eight, which would indicate that all eight samples have at least one read per million for the gene in question. We will drop genes that are known from fewer than one read per million if that situation occurs in three or more of our eight samples.

```
keep<-rowSums(cpm(Shrna_List)>1)>=3
Shrna_List<-Shrna_List[keep,,keep.lib.sizes=FALSE]
rm(keep)
```

In this case, no genes were removed.
Here we apply scale normalization to the shRNA read counts, using the TMM normalization method:

```
Shrna_List<-calcNormFactors(object=Shrna_List,method="TMM")
```

5.1 Differential Counts by the Limma-Trend Method

Limma-trend is a straightforward test for differential counts analysis that works well when all the samples to be analyzed have approximately the same sequencing depth. The yardstick used in the Limma User's Guide is a ratio of the largest library size to the smallest is less than 3-fold:

```
(Depth<-apply(Shrna_List,MARGIN=2,FUN=sum))

## Initial.1 Initial.2 EnzN.1 EnzN.2 EnzN.3 EnzP.1 EnzP.2 EnzP.3
## 22141545 19125653 20202085 19523160 20178264 21474089 20831135 22619107

max(Depth)/min(Depth)

## [1] 1.182658
```

In our case, the ratio of the largest library size to the smallest is 1.182658, so the key assumption of equivalent library sizes is satisfied.

For the Limma-Trend method, counts are converted to logCPM values using edgeR's cpm function. The prior.count is a count to be added to each observation to avoid taking the log of zero, which also reduces the variances of logarithms with low counts:

```
logCPM<-cpm(Shrna_List,log=T,prior.count=3)
```

Here we make a design matrix to inform Limma of our test group (EnzPos) and both control groups (EnzNeg and Initial):

```
FF<-as.factor(c("Initial", "Initial", "EnzNeg", "EnzNeg", "EnzNeg", "EnzPos", "EnzPos", "EnzPos"))
design<-model.matrix(~0+FF)
colnames(design)<-c('EnzNeg', 'EnzPos', 'Initial')
design

##   EnzNeg EnzPos Initial
## 1      0      0      1
## 2      0      0      1
## 3      1      0      0
## 4      1      0      0
## 5      1      0      0
## 6      0      1      0
## 7      0      1      0
## 8      0      1      0
## attr("assign")
## [1] 1 1 1
## attr("contrasts")
## attr("contrasts")$FF
## [1] "contr.treatment"
```

Applying the design matrix to the fitted object:

```
fit<-lmFit(object=logCPM, design=design)
```

Building the contrast matrix:

```
(contrast.matrix<-makeContrasts(contrasts=c("EnzPos-EnzNeg",
                                           "EnzPos-Initial"),
                               levels=design))

##           Contrasts
## Levels      EnzPos-EnzNeg EnzPos-Initial
##   EnzNeg             -1             0
##   EnzPos              1             1
##   Initial              0            -1
```

Applying the contrast matrix:

```
fit2<-contrasts.fit(fit=fit, contrasts=contrast.matrix)
fit2<-eBayes(fit=fit2, trend=T)
```

Top ten significant differences in EnzPos vs EnzNeg:

```
topTable(fit2, coef=1, adjust="BH")

##           logFC AveExpr      t      P.Value      adj.P.Val      B
## ACTB      0.2556190 9.825099  9.167655 1.181564e-09 5.960988e-06 11.464754
## PTGES2 -0.3130965 7.778680 -6.043384 2.144062e-06 4.613290e-03 4.872022
## CD9       0.2578492 7.973403  5.948169 2.743284e-06 4.613290e-03 4.649149
## ACYP2    -0.2386888 7.997590 -5.612229 6.583742e-06 8.303744e-03 3.855391
## RAG2     -0.2374657 8.333637 -5.429277 1.064092e-05 1.060330e-02 3.418886
## ASCC3     0.2125217 7.994139  5.319595 1.420300e-05 1.060330e-02 3.156016
## CARM1    -0.2380458 7.705981 -5.281706 1.569492e-05 1.060330e-02 3.065029
## DEDD     -0.2141019 8.290076 -5.233065 1.784387e-05 1.060330e-02 2.948099
## SLC6A12  0.2202579 8.188632  5.194965 1.973202e-05 1.060330e-02 2.856418
## CA3      -0.2032963 8.123315 -5.094529 2.572968e-05 1.060330e-02 2.614394
```

A positive value in the logFC column indicates more counts in Enz-Pos compared to Enz-Neg, and a negative value indicates more counts in Enz-Neg compared to Enz-Pos.
Top ten significant differences in EnzPos vs Initial:

```
topTable(fit2,coef=2,adjust="BH")
```

##		logFC	AveExpr	t	P.Value	adj.P.Val	B
##	PSMA2	-2.8242340	6.050684	-37.83531	2.232623e-24	1.126358e-20	45.31719
##	EIF3A	-1.2511407	8.688532	-30.29745	6.593115e-22	1.663113e-18	39.99698
##	KIF11	-1.2364350	8.107365	-26.88846	1.364174e-20	2.294086e-17	37.08205
##	RBX1	-0.8690754	9.482671	-25.74060	4.104568e-20	5.176887e-17	36.01104
##	CDK4	-1.4088616	7.203762	-25.06515	8.018847e-20	7.434626e-17	35.35731
##	PSMB2	-1.8156250	6.289746	-24.96797	8.841973e-20	7.434626e-17	35.26177
##	RPL11	-2.0990509	5.816691	-24.45420	1.491014e-19	1.074595e-16	34.75020
##	RPL30	-1.8500293	5.987768	-24.29048	1.764885e-19	1.112980e-16	34.58487
##	TP53INP1	1.0963357	9.039551	24.15189	2.037357e-19	1.142052e-16	34.44403
##	PSMB4	-1.5487073	6.554663	-23.96735	2.469454e-19	1.245840e-16	34.25521

A positive value in the logFC column indicates more counts in Enz-Pos compared to Initial, and a negative value indicates more counts in Initial compared to Enz-Pos.
A given shRNA was considered a “hit” if it showed at least 2-fold abundance decrease in Enz-Pos relative to both Enz-Neg control and initial samples. There are no genes in the first contrast (EnzPos vs. EnzNeg) with a 2-fold abundance decrease, so we tested for any genes with conserved abundance decreases in both sets:

```
Como1<-subset(x=topTable(fit2,coef=1,number=nrow(fit2),adjust="BH",sort.by="logFC"),
              subset=(logFC<=-0.2))
Como2<-subset(x=topTable(fit2,coef=2,number=nrow(fit2),adjust="BH",sort.by="logFC"),
              subset=(logFC<=-0.5))
(Hits<-intersect(x=rownames(Como1),y=rownames(Como2)))

## [1] "TFAP2C" "CAD" "SPDEF" "ITGB4BP" "GABRG2" "CDC37" "PSMD12" "COL5A2"
## [9] "AR" "MAP3K11" "ACAT1"
```

We found 11 genes out of 5045 that satisfy the criteria above to be considered “hits”. The table below shows the log-scaled shRNA counts for the hits:

```
xtable(coefficients(fit)[Hits,c("EnzPos","EnzNeg","Initial")])
```

	EnzPos	EnzNeg	Initial
TFAP2C	6.19	6.49	7.12
CAD	5.48	5.75	6.90
SPDEF	6.07	6.32	7.03
ITGB4BP	6.13	6.36	7.27
GABRG2	7.49	7.72	8.02
CDC37	6.35	6.57	7.06
PSMD12	6.65	6.86	7.68
COL5A2	5.64	5.85	6.60
AR	6.79	7.00	7.70
MAP3K11	6.97	7.18	7.72
ACAT1	6.67	6.88	7.24

Here we create two volcano plots, which are convenient diagrams that show the fold changes versus a measure of statistical significance of the change. This code was modified from the native Limma volcano plot code to highlight only our hits.

```
MyCoefs<-as.matrix(fit2$coef)[,1]
MyLods<-as.matrix(fit2$lods)[,1]
plot(x=MyCoefs,
     y=MyLods,
     xlab="Log Fold Change",
     ylab="Log Odds",
     pch=16,
     cex=0.35,
     main=colnames(coefficients(fit2))[1])
text(x=MyCoefs[Hits],y=MyLods[Hits],labels=Hits,cex=0.8,col="blue")
```

Second volcano plot:

```
MyCoefs<-as.matrix(fit2$coef)[,2]
MyLods<-as.matrix(fit2$lods)[,2]
plot(x=MyCoefs,
     y=MyLods,
     xlab="Log Fold Change",
     ylab="Log Odds",
     pch=16,
     cex=0.35,
     main=colnames(coefficients(fit2))[2])
text(x=MyCoefs[Hits],y=MyLods[Hits],labels=Hits,cex=0.8,col="blue")
```

The shRNA count differences are more striking in the EnzPos *vs.* Initial contrast compared to the Enz-Pos *vs.* EnzNeg contrast.

5.2 Differential Counts with edgeR's Negative Binomial Method

We also calculated counts differences using the negative binomial method implemented in edgeR, and achieved identical results as the above test with Limma's Limma-Trend method.

First we estimated the dispersions:

```
Shrna_List<-estimateDisp(y=Shrna_List,design=design)
```

Then we discovered genes with differential counts:

```
top1<-topTags(exactTest(Shrna_List,pair=c("EnzNeg","EnzPos")),n=length(rownames(Shrna_List)))
Como1<-subset(x=top1[[1]],subset=(logFC<=-0.2))
top2<-topTags(exactTest(Shrna_List,pair=c("Initial","EnzPos")),n=length(rownames(Shrna_List)))
Como2<-subset(x=top2[[1]],subset=(logFC<=-0.5))
(HitsEdgeR<-intersect(x=rownames(Como1),y=rownames(Como2)))

## [1] "GABRG2" "SPDEF" "CAD" "TFAP2C" "PSMD12" "MAP3K11" "AR" "ITGB4BP"
## [9] "ACAT1" "CDC37" "COL5A2"
```

6 Compiling Instructions

This document was created using knitr [4] and $\text{\LaTeX}$ in the RStudio Server (<http://www.rstudio.com/>) integrated development environment (version 0.99.903). Every time the $\text{\LaTeX}$ code is compiled, the code that supports our conclusions is rerun, and in most cases, the results are embedded in this document, facilitating reproducible research. The compile time for this document and its embedded code is approximately nine minutes on our compute server. The .tex and .aux files generated by knitr were compiled to .pdf with pdf_latex and B_BT_EX.

6.1 Build Citation Database

Here we collect the citation information for all packages and databases we used in this report:

```
require(knitr)
suppressWarnings(write_bib(file='Package_citations.bib',tweak=T))
```

6.2 System and Session Information

```
date()

## [1] "Tue Feb 7 14:33:41 2017"

system('rstudio-server version',intern=T)

## [1] "0.99.903"

system('cat /etc/issue',intern=T)

## [1] "Ubuntu 16.04.1 LTS \n \l" ""

system('uname -srv',intern=T)

## [1] "Linux 4.4.0-36-generic #55-Ubuntu SMP Thu Aug 11 18:01:55 UTC 2016"
```

```
toLatex(sessionInfo())
```

- R version 3.3.1 (2016-06-21), x86_64-pc-linux-gnu
- Locale: LC_CTYPE=en_US.UTF-8, LC_NUMERIC=C, LC_TIME=en_US.UTF-8, LC_COLLATE=en_US.UTF-8, LC_MONETARY=en_US.UTF-8, LC_MESSAGES=en_US.UTF-8, LC_PAPER=en_US.UTF-8, LC_NAME=C, LC_ADDRESS=C, LC_TELEPHONE=C, LC_MEASUREMENT=en_US.UTF-8, LC_IDENTIFICATION=C
- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other packages: edgeR 3.14.0, highlight 0.4.7, knitr 1.15.1, limma 3.28.21, plyr 1.8.4, tidyr 0.6.0, xtable 1.8-2
- Loaded via a namespace (and not attached): assertthat 0.1, evaluate 0.10, grid 3.3.1, highr 0.6, lattice 0.20-34, locfit 1.5-9.1, magrittr 1.5, Rcpp 0.12.7, splines 3.3.1, stringi 1.1.2, stringr 1.1.0, tibble 1.2, tools 3.3.1

References

- [1] Andrews,S. FastQC A Quality Control tool for High Throughput Sequence Data.
- [2] Chen,Y., Lun,A., McCarthy,D., Zhou,X., Robinson,M. and Smyth,G. (2016) *edgeR: Empirical Analysis of Digital Gene Expression Data in R*. R package version 3.14.0.
- [3] Smyth,G., Hu,Y., Ritchie,M., Silver,J., Wettenhall,J., McCarthy,D., Wu,D., Shi,W., Phipson,B., Lun,A., Thorne,N., Oshlack,A., de Graaf,C., Chen,Y., Langaas,M., Ferkingstad,E., Davy,M., Pepin,F. and Choi,D. (2016) *limma: Linear Models for Microarray Data*. R package version 3.28.21.
- [4] Xie,Y. (2016) *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.15.1.