SMRT Sequencing of Full-Length Androgen Receptor Isoforms in Prostate Cancer Reveals Previously Hidden Drug Resistant Variants

Tyson A. Clark[1], Manish Kohli[2], Yeung Ho[3], Jamie L. Van Etten[3], Yingming Li[3], Scott M. Dehm[3], David W. Hillman[4], Christine Henzler[5], Rendong Yang[5], Kevin Silverstein[5], Elizabeth Tseng[1], Ting Hon[1], Laura Nolden[1], Liguo Wang[4], and Liewei Wang[6]

[1] PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025
[2] Department of Oncology, Mayo Clinic, 200 1st St. SW, Rochester, MN 55905
[3] Masonic Cancer Center, University of Minnesota, 420 Delaware St. SE, Minneapolis, MN 55455
[4] Division of Biomedical Statistics and Informatics, Department of Health Services, Mayo Clinic, 200 1st St. SW, Rochester, MN 55905
[5] Minnesota Supercomputing Institute, University of Minnesota, 117 Pleasant St. SE, Minneapolis, MN 55455
[6] Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, 200 1st St. SW, Rochester, MN 55905

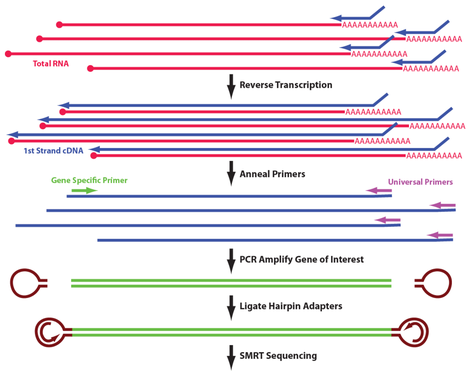
Abstract

Prostate cancer is the most frequently diagnosed male cancer. For prostate cancer that has progressed to an advanced or metastatic stage, androgen deprivation therapy (ADT) is the standard of care. ADT inhibits activity of the androgen receptor (AR), a master regulator transcription factor in normal and cancerous prostate cells. The major limitation of ADT is the development of castration-resistant prostate cancer (CRPC), which is almost invariably due to transcriptional re-activation of the AR. One mechanism of AR transcriptional re-activation is expression of AR-V7, a truncated, constitutively active AR variant (AR-V) arising from alternative AR pre-mRNA splicing. Noteworthy, AR-V7 is being developed as a predictive biomarker of primary resistance to androgen receptor (AR)-targeted therapies in CRPC. Multiple additional AR-V species are expressed in clinical CRPC, but the extent to which these may be co-expressed with AR-V7 or predict resistance is not known.

Here, we utilized long-read sequencing to identify and quantify AR isoforms expressed in CRPC. To unambiguously characterize all AR isoforms, we prepared Iso-Seq libraries via 3' rapid amplification of cDNA ends (RACE) with RNA isolated from prostate cancer cell lines and xenograft tissues using a forward primer anchored in AR exon 1. 3'RACE reactions were subjected to Single Molecule, Real-Time (SMRT) long-read sequencing with a PacBio RS II System.

Our work identified AR-V9 as a truncated isoform that is frequently co-expressed with AR-V7 in CRPC. Mechanistically, our work re-annotated AR-V7 and AR-V9 mRNAs, showing these two species shared a common 3' terminal exon containing separate splice acceptor sites. Using this new information, novel siRNAs and antibodies that could distinguish between AR-V7 and AR-V9 were designed, validated and used to measure the relative expression of these two AR isoforms in CRPC cells with a view to determining the potential of AR-V9 as a predictive biomarker of primary resistance to AR-targeted therapies.

Method Overview



SMRT 3' RACE
(Rapid Amplification of cDNA Ends)

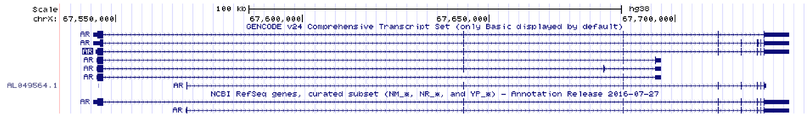
3' or 5' RACE works well for targeted sequencing of cDNAs when the gene of interest includes alternative transcriptional termination or start sites.

Full-length first strand cDNA is generated using an oligo dT primer with a universal amplification sequence appended to the end. The gene of interest is specifically amplified using a custom primer. The double-stranded amplicon can then be converted into a SMRTbell template by ligation of hairpin adapters and sequencing on a PacBio RS II or Sequel System.

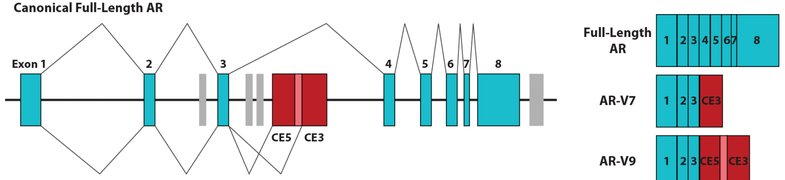
Long sequencing reads from PacBio enable sequencing near full-length RACE amplicons by using a gene-specific primer near the 5' end of the gene.

Alternative Splicing of the Androgen Receptor

Genome Browser View of AR Gene Locus

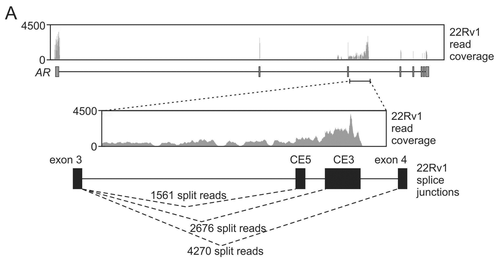


Canonical Full-Length AR



Truncated AR Variants

The canonical full-length version of the Androgen Receptor (AR) includes exons 1-8. However, truncated versions of the receptor protein can be generated by alternative splicing to one or more cryptic exons. AR-V7 arises from contiguous splicing of AR exons 1, 2, 3, and cryptic exon 3 (CE3). The AR-V9 variant utilizes an upstream splice acceptor site incorporating cryptic exon 5 (CE5). CE3 and CE5 are a single alternative terminal exon with two different splice site acceptors. Proteins made from the AR-V isoforms retain the DNA binding domain, but lack the ligand binding domain.



Deep Sequencing of AR RNA in Castration-Resistant Prostate Cancer.
(A) RNA-Seq read coverage along discrete regions of the AR gene in the 22Rv1 cell line. Split RNA-Seq reads spanning the exon 3/CE5, 3/CE3, and 3/4 splice junctions were quantified to infer expression of AR-V9, AR-V7, and full-length AR, respectively. However, in short-read sequencing data, it is difficult to infer contiguous full-length mRNA isoforms. This makes it a challenge to understand co-expression of multiple variants.

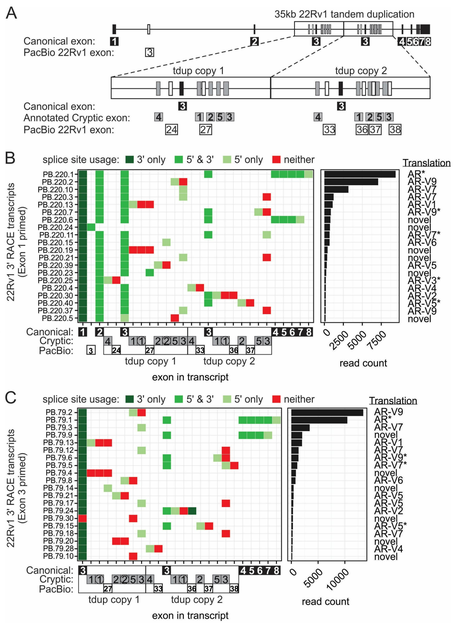
Full-Length Transcript Sequencing Identifies Novel Isoforms

Single Molecule Real-Time (SMRT) Sequencing of AR Isoforms in Castration-Resistant Prostate Cancer

(A) Schematic of AR gene structure in the 22Rv1 cell line, which harbors a 35 kb AR intragenic tandem duplication (tdup). Discrete exons expressed in AR mRNAs sequencing by PacBio SMRT Sequencing are illustrated. Black exons represent AR exons 1-8 (canonical exons), gray exons represent known AR cryptic exons (annotated cryptic exons), and white exons represent new AR exons discovered by SMRT Sequencing (PacBio 22Rv1 exons).

(B) Exon composition and quantification of the 20 most abundant AR transcripts isolated from 22Rv1 cells by 3' rapid amplification of cDNA ends (RACE) using a forward primer anchored in AR exon 1. Individual pixels represent discrete exons contained in individual AR transcripts. Pixel colors indicate whether that exon was spliced via both annotated 5' and 3' splice sites, the annotated 5' splice site only, the annotated 3' splice site only, or neither the annotated 5' nor 3' splice sites. Read counts represent the number of single molecule transcripts that matched the indicated splicing pattern. AR transcripts were inspected manually for predicted translation, and assigned names based on a previous classification system. AR transcripts that had not been identified previously were classified as novel. Asterisks denote transcripts harboring tandem copies of AR exon 3.

(C) Exon composition and quantification of the 20 most abundant AR transcripts from 22Rv1 cells by 3' RACE using a forward primer anchored in AR exon 3. Data are visualized as described above.



Summary and Resources

- 3' RACE combined with PacBio SMRT Sequencing enables targeted full-length isoform identification
- SMRT Sequencing of AR transcripts provides unambiguous full-length mRNA information
- Alternatively spliced AR transcripts that incorporate cryptic exons result in proteins that are likely to be truncated
- AR-V9 mRNA is frequently co-expressed with AR-V7 and both share a common 3' terminal exon
- The newly-annotated features of the human AR gene are important for the design and interpretation of targeted assays for developing AR-Vs as biomarkers in prostate cancer

More information on full-length transcript sequencing (Iso-Seq Application) can be found on the PacBio website: http://pacb.com/isoseq