ORIGINAL PAPER

Identification of Androgen Receptor Splice Variant Transcripts in Breast Cancer Cell Lines and Human Tissues

**Dong Gui Hu · Theresa E. Hickey · Connie Irvine ·
Dhilushi Dodampege Wijayakumara · Lu Lu ·
Wayne D. Tilley · Luke A. Selth · Peter I. Mackenzie**

Abstract The androgen receptor (AR) is widely expressed in human tissues and has biological function in many male and female organs. In particular, the AR plays a critical role in the biology and pathology of the prostate gland. AR activity inhibits breast growth and has pleiotropic actions in breast cancer that are subtype-dependent. Expression of AR splice variants (ARVs) and their role in prostate carcinogenesis has been elucidated in recent studies. We hypothesised that ARVs are also expressed in breast cancers and other hormone sensitive tissues. Herein, the expression of five previously identified ARV transcripts with documented transcriptional capacity (AR-V1, -V3, -V4, -V7, and -V9) was examined in 6 breast (MFM223, MDA-MB-453, MDA-MB-231, ZR75.1, MCF-7, T47D), two prostate (VCaP, LNCaP), and one liver (HepG2) cancer cell lines, a human embryonic kidney cell line (HEK293), and a panel of RNAs representing 21 different human tissues. Four ARVs (V1, V3, V7, V9) were detected to some degree in almost all cell lines and tissues. In addition, four novel ARVs containing a cryptic exon 9 (CE9) were detected in MDA-MB-453 and VCaP cells. Sequencing of ARV amplicons revealed a single nucleotide substitution within CE3 in lung and placental tissue samples that could be translated as an Ile (ATT)>Val (GTT) substitution in the AR-V7 variant protein. Collectively, these data provides insight into the potential complexity of AR transcriptional splicing events in breast cancer cell lines and diverse human tissues, thereby establishing a rationale for further exploration of ARVs in breast cancer and other human pathologies.

Abbreviations

ADT	Androgen deprivation therapy
AR	Androgen receptor
AR-FL	Full length AR
ARV	Androgen receptor splicing variant
CE	Cryptic exon
CRPC	Castration-resistant prostate cancer
DBD	DNA binding domain
ER	Estrogen receptor
FBS	Fetal bovine serum
LAR	Luminal androgen receptor
LBD	Ligand-binding domain
NLS	Nuclear localization signal
NTD	N-terminal domain
PSA	Prostate-specific antigen

D. G. Hu (✉) · D. D. Wijayakumara · L. Lu · P. I. Mackenzie
Department of Clinical Pharmacology, Flinders Medical Centre,
Flinders University School of Medicine, Bedford Park, SA 5042,
Australia
e-mail: dongui.hu@flinders.edu.au

T. E. Hickey · C. Irvine · W. D. Tilley (✉) · L. A. Selth
Dame Roma Mitchell Cancer Research Laboratories and Adelaide
Prostate Cancer Research Centre, School of Medicine, The
University of Adelaide, DX Number 650 801, Adelaide, SA 5005,
Australia
e-mail: wayne.tilley@adelaide.edu.au

Introduction

The androgen receptor (AR) is a sex steroid hormone receptor that directly mediates the actions of androgen hormones. Development of the male phenotype and adult reproductive function is dependent on a functional AR [1], while the role of AR in establishing a female phenotype and regulating adult reproductive function appears to be more modulatory than essential [2]. The AR is widely expressed in many male and female tissues, but the molecular biology of this sex steroid

receptor has largely been characterized in studies of the prostate gland, where the AR has critical functions in regulating its development and normal physiology as well as mediating pathology, including cancer development and progression [3]. Indeed, AR is considered the key oncogenic driver at all stages of prostate cancer. Accordingly, androgen deprivation therapy (ADT), achieved via blockage of androgen production through castration and/or antagonism of AR activity by anti-androgens, forms the cornerstone of endocrine therapy for this disease [4]. Despite an initial response to ADT, men with prostate cancer invariably progress to an incurable state called castration-resistant prostate cancer (CRPC), thought to occur via therapy-driven adaptive mechanisms that sustain or resurrect an active AR signaling axis in prostate cancer cells [5]. In recent years, increased expression of variant transcripts from the AR gene (ARVs) has been shown to represent one such adaptive mechanism [6].

The AR gene is found at Xq11-13 and has 8 canonical exons that encode a modular transcription factor consisting of an N-terminal transactivation domain (NTD, encoded by exon 1), a central DNA binding domain (DBD, exons 2-3), a hinge region harboring the second half of the bipartite nuclear localization signal (NLS, exon 4), and a C-terminal ligand-binding domain (LBD, exons 5-8) [7, 8]. In addition, at least 7 other so-called "cryptic" exons (CE1-5, CE1b, and CE9) have been identified in the AR gene [9–13]. These cryptic exons can be incorporated into AR transcripts by altered splicing events [14] or as a result of genomic rearrangements [15, 16], yielding at least 14 variant AR transcripts (termed AR-V1 to AR-V14) [10]. As reviewed recently, the majority of ARVs identified to date in prostate cancer cells are generated through splicing of exon 3 to a downstream cryptic exon, producing variant proteins with the NTD/DBD component attached to a C-terminal unique peptide of varying length, thereby splicing out the majority of the LBD [10]. Most ARVs have been shown to be constitutively (AR-V3, -V4, -V7, and -V12) [9, 10, 12, 17] or conditionally (AR-V1 and -V9) [10, 12] active androgen-independent transcription factors in a prostate cancer context. Pre-clinical and clinical studies have supported the concept that ARVs are biologically relevant in prostate cancer, particularly in the genesis of a CRPC [11, 12, 17–20].

In addition to prostate cancer, AR plays a role in other androgen-sensitive malignancies including those of the breast, bladder, kidney, lung, and liver [21, 22]. In particular, the role of AR in breast carcinogenesis has been the topic of increasing interest over the past decade. While it is clear that AR is expressed in the large majority of breast cancers, controversy exists over its role, which appears to vary depending on the molecular subtype [23]. To date, studies examining the expression of ARVs in human tissues or malignancies other than the prostate have been limited. An ARV called AR45 (an N-terminal truncated variant) has been identified in various human tissues, most strongly in the heart, and multiple ARVs

with a deletion of exon 2, 3, or 4 have been identified in human testis [13, 24]. An exon 3-deleted ARV has also been observed in breast cancers [25]. Whether C-terminal truncated ARVs found recently in prostate cancer are also expressed in breast cancers and other human tissues remains unexplored.

In this study, evidence is presented for expression of multiple known and novel ARV transcripts in a panel of breast cancer cell lines and human tissues. These observations suggest that ARVs merit further investigation in steroid sensitive tissues and malignancies, particularly in breast cancer.

Materials and Methods

Human Total RNA Survey Panel

A total RNA panel of 21 human tissues (prostate, liver, colon, spleen, lung, testis, kidney, placenta, bladder, brain, adipose, ovary, cervix, heart, skeletal muscle, small intestine, thyroid, thymus, esophagus, trachea, and breast) was purchased from Ambion (Applied Biosystems). As stated in the manufacturer's instructions, each total RNA sample was comprised of at least 3 tissue donors and prepared from tissues that had been frozen or stored in RNAlater® until processing. Integrity of the RNA was verified by the supplier using capillary electrophoresis on an Agilent2100® bioanalyzer. We used 1 μg of RNA per sample to generate cDNA, as described in detail below.

Genomic DNA Samples

Twenty one genomic DNA samples for SNP genotyping were previously isolated from human liver tissues from Caucasian donors obtained from the Liver Bank of the Department of Clinical Pharmacology, Flinders Medical Centre [26, 27]. Details of the donors have been previously published [28].

Cell Culture, RNA Extraction, and Reverse Transcription

Ten human cell lines were used in this study including transformed human embryonic kidney cells (HEK293), liver cancer cells (HepG2), breast cancer cells (MFM223, MDA-MB-453, MDA-MB-231, ZR75.1, MCF-7, T47D), and prostate cancer cells (VCaP, LNCaP). The MFM223 cell line was from the European Collection of Cell Cultures via Sigma-Aldrich. The remaining nine cell lines were purchased from American Type Culture Collection (ATCC). All cell lines underwent verification by short-tandem repeat profiling in 2013 by CellBank Australia. VCaP and HepG2 cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10 % fetal bovine serum (FBS). Five cell lines (MDA-MB-231, MDA-MB-453, ZR75.1, MCF-7, and HEK293) were maintained in RPMI-1640 containing 5 % FBS. T47D

cells were cultured in phenol red-free RPMI-1640 with 5 % FBS. MFM223 cells were cultured in Minimum Essential Medium Eagle (MEME, Sigma) containing 10 % FBS, 1xITS (a mixture of recombinant human insulin, human transferrin, and sodium selenite), and 2 mM L-glutamine. Cells were plated in 6-well plates in triplicate for each cell line and cultured for 3–5 days until cells reached 90 % confluence. Total RNA was extracted from the cells of each well using RNeasy Mini Kits (QIAGEN, Valencia, CA) according to the manufacturer's instructions. Reverse transcription (RT) was carried out using reagents from Invitrogen as previously reported [29]. Briefly, total RNA (1 μg) was treated with DNase I at room temperature for 15 min and then reverse-transcribed at 50 °C for 50 min in a final volume of 20 μl of 1xfirst strand buffer (50 mM Tris-HCl, pH 8.0, 75 mM KCl, and 3 mM MgCl2) containing random hexamer primers, 50 units of RNaseOut recombinant ribonuclease inhibitor, and 50 units of Superscript III. The resulting cDNA was diluted into 80 μl of RNase-free water. 3 μl of each diluted sample (~60 ng) was used as template in quantitative real-time PCR as described below.

To examine the effect of androgen on ARV expression, MDA-MB-453 cells were seeded in six-well plates and cultured in RPMI containing 5 % FBS. 48 h later, medium was replaced with fresh RPMI medium containing 5 % FBS (three wells, RPMI) or phenol red-free RPMI containing 5 % dextran-coated charcoal (DCC) stripped FBS (six wells, DCC). After a further 48 h, RPMI-cultured cells were refreshed with RPMI medium containing 5 % FBS, and the DCC-cultured cells were treated either with vehicle (0.1 % ethanol) or 1 nM dihydrotestosterone (DHT). 24 h later, RNA was harvested from cells and subjected to RT-qPCR to quantify target transcripts including 18S rRNA, ARV1, AR-V3, AR-V7, and AR-V9 as described below.

Quantitative Real-Time PCR (qPCR)

Of the 14 ARVs reviewed by Hu et al [10], five are generated by the splicing of a cryptic exon to exon 2 (AR-V3) or exon 3 (AR-V1, -V4, -V7, and -V9) of the *AR* gene and all of these have demonstrated hormone independent transcriptional capacity. In contrast, AR45, an N-terminally truncated ARV generated through splicing of cryptic exon 1b to exon 2 of the *AR* gene is a hormone dependent ARV that can inhibit AR-FL activity in prostate cancer cells or independently activate transcription when co-expressed with certain transcriptional co-activators [13]. Due to demonstrated functional activity in a hormone regulated model system, specific primers (Table 1) were designed to quantify these six ARV transcripts, along with AR-FL and β-actin, by quantitative real-time PCR (qPCR) in the cell lines and human tissues. The amplified region of each target and the sizes of the amplicons are shown in Fig. 1a. PCR products representing each amplicon were

first generated using VCaP cDNA and the identities of the resultant PCR products were validated by sequencing. To enable generation of standard calibration curves for the qPCR analyses, the amplicon of each target transcript was cloned into the PCR-blunt vector (Invitrogen) according to manufacturer's protocol. Four serial 10-fold dilutions containing known copy numbers (e.g., 6,000, 600, 60, and 6) of target-containing PCR-blunt vectors were included in each qPCR run in order to calculate copy numbers of target transcripts in the experimental samples being amplified in the same run. Representative standard curves for four ARVs (V1, V3, V7, and V9), AR45, and AR-FL are given in Supplementary Fig. 1. As previously reported, the qPCR product of 18S rRNA was cloned and quantified in a similar manner during each run [29]. qPCR was performed using a RotorGene 3000 instrument (Corbett Research, NSW, Australia) in 20 μl of 1x QIAGEN QuantiTect SYBR Green PCR master mix containing 3 μl (~60 ng) of cDNA sample and target-specific forward and reverse primers (500 nM each). The amplification conditions consisted of an initial activation step of 95 °C for 15 min, and 40 cycles of 95 °C for 10 s, 56 °C to 60 °C for 15 s, and 72 °C for 20 s. Data were obtained during the 72 °C extension phase of each cycle and analyzed by the program RotorGene 6.1 (Corbett Research). Copy numbers of 18S rRNA transcripts were used to normalize the amount of total RNA amplified in each reaction. Each sample was amplified in duplicate and the resultant mean values were used for analysis. The target transcript expression levels were expressed as copy numbers of each target transcript per 10^9 copies of 18S rRNA in the same samples. All qPCRs were repeated at least twice and similar results were observed.

Cloning of Exons 6-9 of the *AR* cDNA

PCR using Phusion® High-Fidelity DNA polymerase was conducted to amplify the region encompassing exons 6-9 of the *AR* using cDNAs generated from a breast and prostate cancer cell line (MDA-MB-453 and VCaP, respectively) using the forward primer at exon 6 (5′-ATGCACAAGTCCCGGA TGTA-3′) and a previously reported reverse primer P2(R) located within the cryptic exon 9 [10]. Resultant PCR products were subsequently cloned into the Invitrogen TA vector (pCR®2.1). Inserts were bidirectionally sequenced using primers T7 and Sp6 (Invitrogen).

Results

Identification of Androgen Receptor Splice Variants in Breast Cancer Cell Lines

AR-FL transcript copy number was very high in the prostate cancer cell lines (VCaP>LNCaP), moderate in the MDA-MB-

Table 1 Primer sequences

Targets	Forward primer (5′to 3′)	Reverse primer (5′to 3′)	Amplificon size (bp)
AR-V1	TGCGCCAGCAGAAATGATTG	GGTTTTCAAATGCAGCCAGG	250
AR-V3	AGACGAAGCTTCTGGGTGT	CATGCAGTATGGCTTGGGGT	139
AR-V4	TGCGCCAGCAGAAATGATTG	AGTATGGCTTGGGGTTAGT	157
AR-V7	TGCGCCAGCAGAAATGATTG	ACCTAAGCTCCTAAGCCTCA	245
AR-V9	TGCGCCAGCAGAAATGATTG	CTGCGTGTTTTTCCCAGT	192
AR45	GCTTTGTACCGGACTTTGTA	GTGGAAAGTAATAGTCAATGG	147
AR-FL	ACATCCTGCTCAAGACGCTTCTACC	CACTTGCACAGAGATGATCTCTGCC	162

453, MFM223, MCF7, ZR75.1 and T47D breast cancer cells (listed from highest to lowest copy number), and very low in the HepG2 liver cancer cells, HEK-293 and MDA-MB-231 breast cancer cells. All breast and prostate cancer cell lines examined expressed AR-V1, AR-V3, AR-V7, AR-V9 and AR45 to some degree (Fig. 1b). ARVs were most abundant in VCaP prostate cancer cells, with expression levels being 4-8 fold (AR-V3, -V7, -V9, AR45) or 80 fold (AR-V1) higher than those observed in LNCaP cells. Compared to LNCaP cells, the breast cancer cell lines generally had a higher copy number of AR-V1 and lower copy numbers of AR-V7, -V9 and AR45. Interestingly, two of the breast cancer cell lines, MDA-MB-231 and ZR75.1, had higher copy numbers of AR-V3 than LNCaP cells. This is surprising because not only do the breast cancer cell lines have considerably lower levels of AR-FL than LNCaP cells, the MDA-MB-231 cells in particular have such low AR-FL expression they are often considered an AR negative breast cancer cell line. Another notable comparative difference among the six breast cancer cell lines is the distinctly higher copy number of AR45 in MDA-MB-453 cells compared to all other lines. When comparing relative ARV to AR-FL copy numbers in each cell line (Fig. 1c), cell type-specific profiles are evident. As expected, the ARV/AR-FL expression ratios across the 10 cell lines are generally less than or approximately equal to 5 %. However, three cell lines (MDA-MB-231, HepG2, and HEK293) that possessed the lowest AR-FL copy numbers had an ARV/AR-FL ratio of >5 % for multiple ARVs. Remarkably, AR-V3 transcript numbers were approximately 75 % as high as those of AR-FL in the MDA-MB-231 cell line.

A recent study with VCaP Xenografts in mice showed a significant increase in ARV expression just 2 days postcastration, and that this increase was completely abrogated by testosterone replacement [30]. We demonstrated that dihydrotestosterone (DHT) negatively regulated ARV expression in the MDA-MB-453 cell line in a similar manner. As shown in Supplementary Fig. 2 , depletion of steroid hormones (e.g., androgens) enhanced the expression of four ARVs (V1, V3, V7, and V9), and this increased expression was completely abolished upon exposure to DHT.

Androgen Receptor Variants in a Panel of Human Tissues

A panel of 21 human tissues were interrogated for expression of AR-FL and ARVs that were detected in the panel of cell lines. Figure 2a shows the copy numbers of each target transcript relative to 10^9 copies of 18S rRNA in the same tissue sample. AR-FL and the five ARV transcripts were detected in almost all tissues, the exceptions being AR-V3 in spleen and trachea and AR-V7 in the thyroid gland. As expected, steroid-sensitive tissues (prostate, testis, adipose, ovary, cervix and breast) had higher copies of AR-FL transcripts compared to other tissues. Among the 5 ARVs, AR45 was generally the most abundant and AR-V3 the least. Heart tissue had the highest number of AR45 transcripts, consistent with a previous report [13]. The other three ARVs (AR-V1, -V7, and V9) were expressed in tissues at equivalent or 2-3 fold lower levels compared to the prostate. As shown in Fig. 2b, the ARV/AR-FL expression ratios across the primary tissue samples were generally less than 2.5 %; however, AR45 was expressed at ~5 % of AR-FL in some tissues (spleen, placenta, small intestine) and approaching 20 % in the heart.

Novel and Known Single Nucleotide Polymorphisms in *AR* Cryptic Exon Coding Sequences

Sequencing of the ARV qPCR products to verify the identities of target amplicons identified an A>G substitution in the cryptic exon CE3 coding sequence that results in an Ile (ATT)>Val (GTT) substitution at codon 5 of the AR-V7-specific C-terminal peptide. This substitution was seen in two tissues (lung and placenta). Both the sense (Fig. 3a) and antisense (Fig. 3b) strands of qPCR products from these two tissues were sequenced to confirm the substitution, ruling it out as a PCR artefact and suggesting that it may represent a novel single nucleotide polymorphism (SNP). According to the supplier's specifications (Ambion, Applied Biosystems), the lung RNA sample was pooled from 2 male and 1 female subjects, and the placental RNA sample was pooled from 3 female subjects. DNA samples were not available to examine this substitution among the contributing individuals. To

Fig. 1 Transcripts of androgen receptor splice variants were detected in breast cancer cell lines. Total RNA was extracted from 10 human cell lines, including 6 breast cancer cell lines (MFM223, MDA-MB-453, MDA-MB-231, ZR75.1, MCF-7, and T47D), 2 prostate cancer cell lines (LNCaP and VCaP), a liver carcinoma cell line (HepG2), and an embryonic kidney cell line (HEK293). RNA was then reverse-transcribed to generate cDNA, followed by quantitative real-time PCR to calculate the copy number of target transcripts as described under *Material and Methods*. **a**: the transcript structures of the wild type AR (AR-FL) and variant ARVs (V1, V3, V7, V9, and AR45) and the regions amplified for transcript quantitation. Shown are qPCR products of target transcripts of 10 cell lines on ethidium bromide-stained agarose gels (2 %). qPCR products of each target were cloned into the PCR blunt vector (Invitrogen) and the resultant plasmids were seeded as positive controls. The negative controls contained water instead of cDNA template. The DNA maker was the 100 bp DNA ladder (New England Biolabs). b: Graphical presentation of the calculated copy numbers of target transcripts per 10^9 copies of 18S rRNA transcripts in 10 cell lines. Data shown are means ± SD of three independent experiments performed in triplicate. **c**: Graphical presentation of the expression ratios of ARV to AR-FL across the 10 human cell lines calculated using the respective means as shown in b
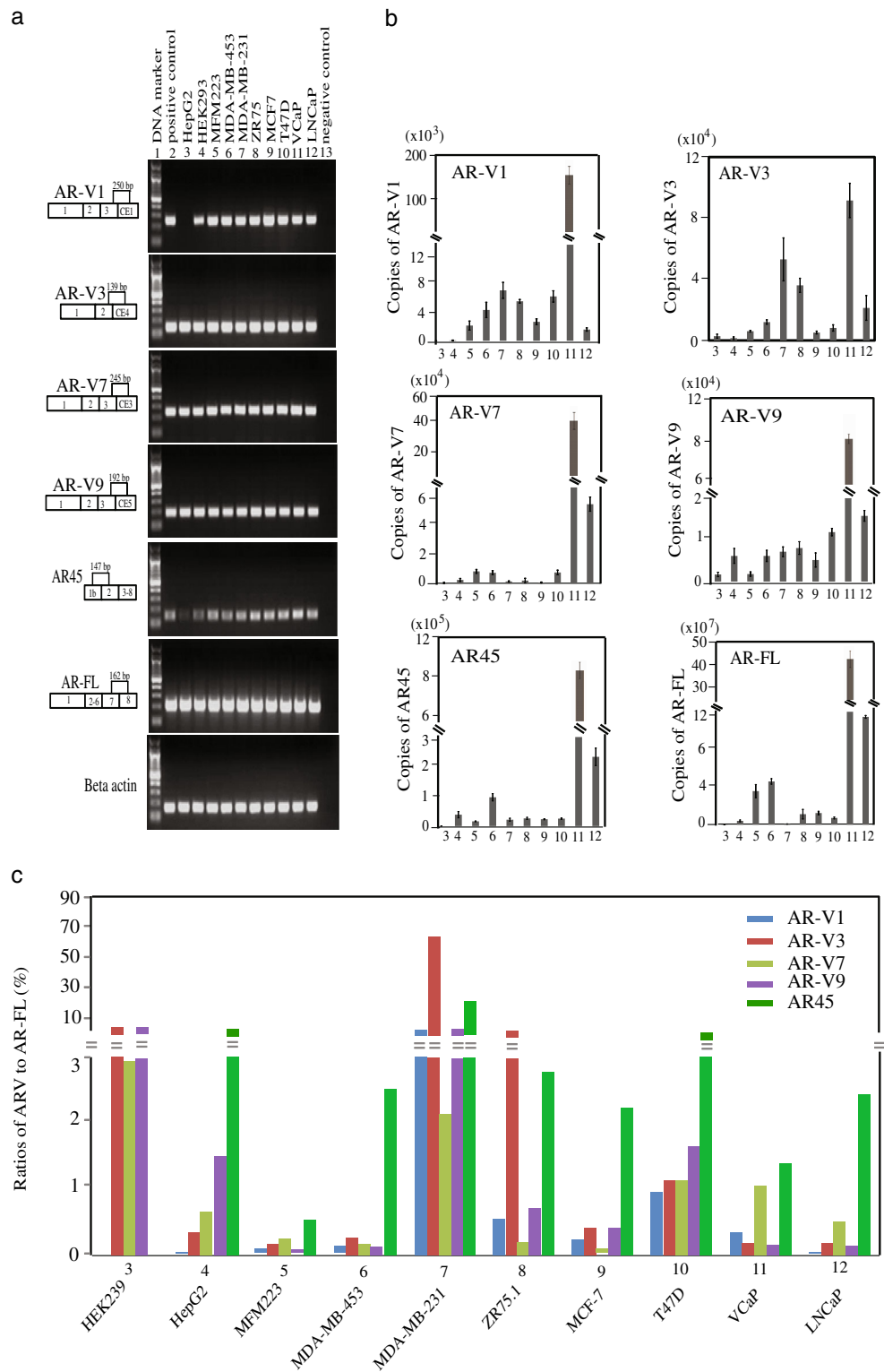
determine the relative frequencies of the two CE3 alleles, the region covering this substitution was amplified from genomic DNA samples of 21 unrelated human subjects and subjected to sequencing. We only found the previously reported CE3-G

allele in this population (data not shown). In addition, CE4 has been reported to code either an 11-amino acid or 53-amino acid peptide [9, 11] depending on the presence of a stop codon (TAA) at codon 12 of the CE4-specific peptide (Fig. 3c). Our

Fig. 2 Transcripts of androgen receptor splice variants were detected in a human tissue panel. Total RNA samples of 21 human tissues were purchased from Ambion (Applied Biosystems), including prostate (1), liver (2), colon (3), spleen (4), lung (5), testis (6), kidney (7), placenta (8), bladder (9), brain (10), adipose (11), ovary (12), cervix (13), heart (14), skeletal muscle (15), small intestine (16), thyroid (17), thymus (18), esophagus (19), trachea (20), and breast (21). cDNAs were generated from these RNA samples, followed by quantitative real-time PCR to calculate the copy numbers of six target transcripts (AR-FL, AR-V1, -V3, -V7, -V9, and AR45) as described under Materials and Methods. **a**: Graphical presentation of the calculated copy numbers of each target transcript per 10^9 copies of 18S rRNA transcripts in 21 human tissues. Data shown are means ± SD of three independent experiments performed in triplicate. **b**: Graphical presentation of the ARV/AR-FL expression ratios across the tissues calculated using the respective means as shown in a
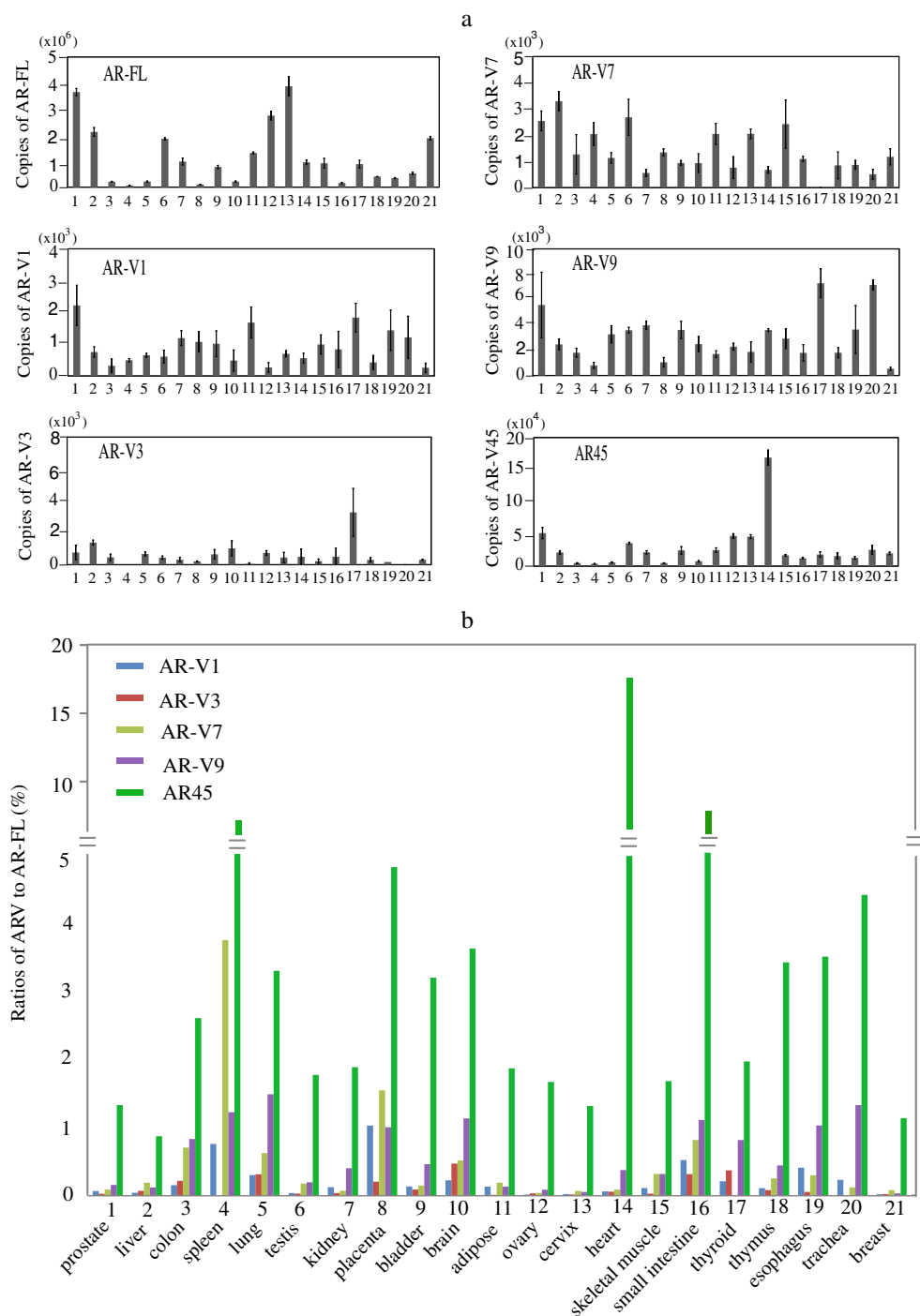
results showed that CE4 had the TAA insertion, which matches the NCBI reference sequence (NT_0011669.17), in all cell lines and tissues.

Encouraged by the finding of the novel substitution, the NCBI dbSNP database was searched for other novel substitutions in the CEs of the *AR* gene and four CEs were discovered to have one or two SNPs within their coding sequences. The SNPs at CE1 (rs1337075C>T) and CE5 (rs138753060A>G) are synonymous substitutions, whereas four SNPs result in

single amino acid substitutions at ARV-specific C-terminal peptides (Table 2). The deduced ARV-specific C-terminal peptides encoded by these SNPs are summarized in Fig. 3c.

Identification of Novel Exon 9-Containing Splicing Variants

Hu et al recently discovered a cryptic exon (termed exon 9) in the 3′ untranslated region (UTR) of the *AR* gene and three exon 9-containing ARVs (V12, V13, and V14) [10]. To

Fig. 3 Single nucleotide substitutions result in single amino acid substitutions in the AR variant-specific C-terminal peptides. **a/b**: Sequencing of the AR-V7-specific qPCR products amplified from cDNAs of two tissues (lung and placenta) identified a same A>G substitution at the CE3 coding sequence that leads to an Ile (ATT)>Val (GTT) substitution at codon 5 of the AR-V7-specific C-terminal peptide. Shown are the sense (**a**) and antisense (**b**) chromatogram sequencing results from the lung tissue cDNA showing the A>G substitution. Genomic coordinates are given according to the NCBI36/HG18 release. The position of the SNP rs370756473G>T is shown with an arrow. Of note, this SNP has not yet been deposited in the dbSNP database but it is currently annotated at the GRCh37.p10 Primary Assembly Reference sequence (NT_011669.17). **c**: single nucleotide polymorphims within the *AR* cryptic exons that are deposited at the NCBI dbSNP database lead to single amino acid substitutions in ARV-specific C-terminal peptides. Shown are both the previously reported ARVs (indicated by a star) encoded by the prevalent alleles and the predicted variant ARVs encoded by the minor alleles. The last 4 amino acids of the canonical exons at the splicing junctions and the amino acids at the substitution positions of the C-terminal variant-specific peptides are in bold
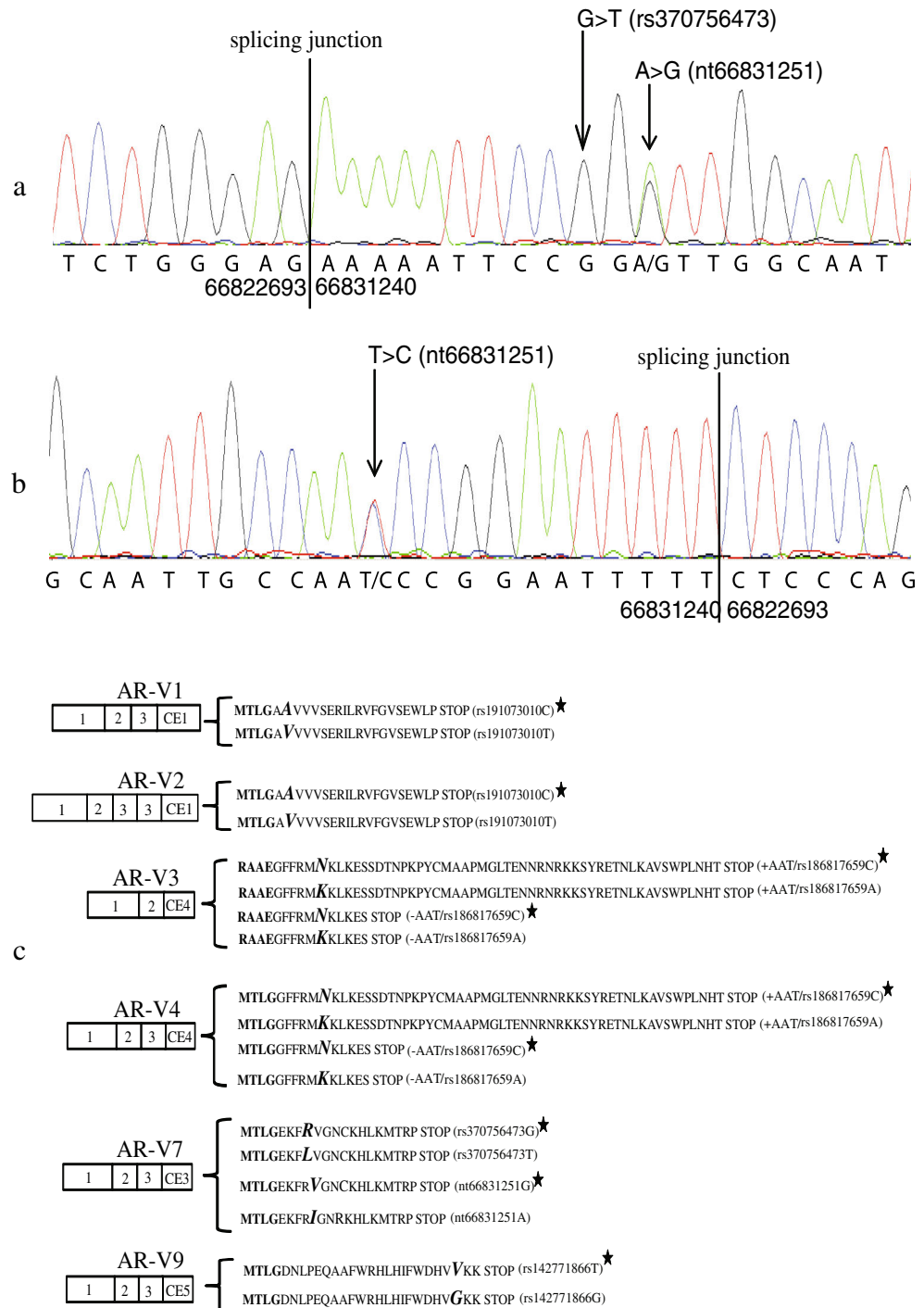
determine whether these three ARVs are expressed in breast cancer cells, a forward primer at exon 6 was paired with a reverse primer P2(R) at exon 9 [11] to amplify the 6510-bp region encompassing exons 6-9 from MDA-MB-453- and VCaP-derived cDNA. Surprisingly, the results consistently revealed amplicons that were smaller than the expected 6510-bp PCR product. As shown in Fig. 4a, both cell lines expressed three amplicons of approximately 1,500, 1,400, and 1,450 bp. VCaP cells also uniquely expressed an amplicon around 3,500 bp in size. The PCR products were cloned and ~100 individual clones were sequenced to determine their identity. This analysis revealed four novel exon 9-containing ARVs. We designated them AR-V15, AR-V16, AR-V17, and AR-V18 in accordance with the nomenclature proposed by Hu et al [10]. As shown in Fig. 4c, according to genomic coordinates of the NCBI36/gh18 release, AR-V15, -V16, -

Table 2 Known and novel single nucleotide polymorphisms at cryptic exons of androgen receptor splice variants
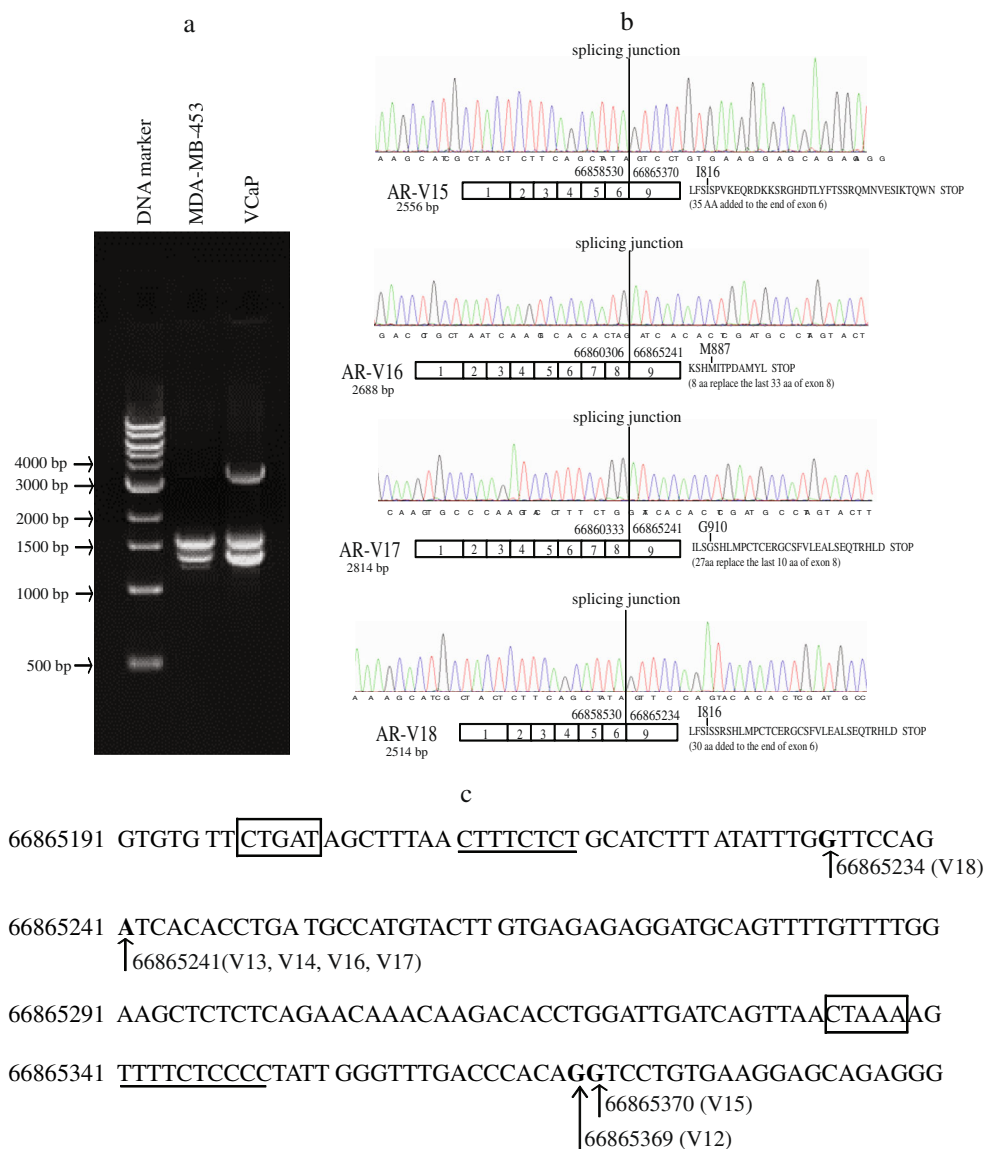
SNP	CE	ARV	Amino acid substitution
rs191073010C>T	CE1	V1/V2	Ala (GCT)>Val (GTT)
rs370756473G>T	CE3	V7	Arg (CGG)>Leu (CTG)
rs186817659A>C	CE4	V3/V4	Asn (AAC)>Lys (AAA)
rs142771866G>T	CE5	V9	Val (GTT)>Gly (GGT)
nt66831251A>G	CE3	V7	Ile (ATT)>Val (GTT)[a]

SNP single nucleotide polymorphism; *CE* cryptic exon; *ARV* androgen receptor splice variant

[a] the SNP identified in this study

V17, and -V18 had novel splice junctions of $A^{66858530}/G^{66865370}$, $G^{66860306}/A^{66865241}$, $G^{66860333}/A^{66865241}$, and $A^{66858530}/G^{66865234}$, respectively. All 4 novel ARVs were detected in MDA-MB-453 cells, whereas three (V15, V16, and V18) were detected in VCaP cells. We observed the AR-V13-specific splicing junction in 14 clones (8 in VCaP and 6 in MDA-MB-453), and the AR-V16-unique splicing junction in 15 clones (8 in VCaP and 7 in MDA-MB-453). The predicted transcript structures of the four novel exon 9-contaning ARVs and their C-terminal unique peptide sequences encoded by exon 9 are shown in Fig. 4b.

Discussion

The AR is widely expressed in human tissues and has pathological implications for many human non-cancerous and cancerous diseases in both men and women [1, 21, 22]. Since the

Fig. 4 Discovery of unreported splicing junctions predicts four novel exon 9-containing AR variants. **a**: PCR amplification of the region encompassing exons 6-9 from cDNAs of MDA-MB-453 and VCaP gave multiple products smaller than the expected size of 6,510 bp as shown on an ethidium bromide-stained agarose gel. **b**: four novel splicing junctions were observed following cloning of the PCR products into the Invitrogen TA vector and sequencing of the resultant clones. Shown are the chromatogram sequencing results containing the novel splicing junctions (genomic coordinates are given according to the NCBI36/HG18 release), the structures and sizes of the four predicted transcripts and their C-terminal unique peptide sequences encoded by exon 9. **c**: exon 9 harbours four cryptic 3′ splice sites. Shown are the 150 bp-genomic sequence of the *AR* gene that contains four cryptic 3′ splice sites (arrowed). The genomic coordinates of the sequence are given at the left and the cryptic splice sites are in bold with branch sites boxed and polypyrimidine tracts underlined

66865191 GTGTG TT CTGAT AGCTTTAA CTTTCTCT GCATCTTT ATATTTG**G**TTCCAG
66865234 (V18)

66865241 **A**TCACACCTGA TGCCATGTACTT GTGAGAGAGGATGCAGTTTTGTTTTGG
66865241(V13, V14, V16, V17)

66865291 AAGCTCTCTCAGAACAAACAAGACACCTGGATTGATCAGTTAA CTAAA AG

66865341 TTTTCTCCCCTATT GGGTTTGACCCACA**GG**TCCTGTGAAGGAGCAGAGGG
66865370 (V15)
66865369 (V12)

AR gene is on the X chromosome and therefore only exists as a single copy per cell in men, functional mutations or genomic variants of the AR generally exert phenotypic consequences. A compendium of genomic anomalies and their phenotypic consequences in men is available on the Androgen Receptor Gene Mutations public database [31]. Among these, mutations in the AR gene that result in abnormal splicing events have long been known to cause some cases of androgen insensitivity syndrome (AIS) [32–34]. However, the association between a disease state and alternative splicing from the wild type AR gene or aberrant splicing due to genomic rearrangements have been very recent discoveries and so far confined to the field of prostate cancer, where the AR has an indisputable role as an oncogene [14, 15, 19, 35]. The AR splice variants expressed in advanced prostate cancer are largely hormone independent transcription factors that contribute to the maintenance of AR signalling in an androgen-deplete environment and are resistant to conventional anti-androgenic drugs as they lack a ligand binding domain, the docking site for these drugs. While the role of AR in breast cancer is still being defined, and AR is likely to exert tumour suppressive as well as oncogenic effects depending on the molecular subtype [23], several clinical trials employing anti-androgen therapy in the treatment of women with advanced breast cancer are underway (NCT00468715, NCT01597193, NCT00755885). In the current study, we provide evidence for the first time that C-terminal truncated ARVs recently discovered in prostate cancer cells are also expressed in breast cancer cell lines and in multiple human tissues. Importantly, if expression of these ARVs is demonstrated at the protein level, in particular in response to AR targeted therapies as has been reported in prostate cancer [35], expression of ARVs may have relevance for women with breast cancer currently being treated with similar AR-targeted therapies as in the clinical trials cited above.

One of the significant discoveries to arise from the EN-CODE project [36] is that the large majority of human genes exhibit some degree of alternative splicing; up to 25 splice variants commonly arise from a single gene in addition to the wild type transcript. The splice variants are generally expressed in a pattern whereby 2-3 variant transcripts are dominant, followed by a wide array of variant transcripts that are expressed at very low copy numbers. Our observations of the pattern of ARV transcript expression in breast cancer cell lines and human tissues are consistent with the ENCODE results. Generally, one or two ARVs were expressed at levels of <5.0 % compared to that of the wild type AR transcript. This data is also consistent with observations of ARV expression in non-cancerous prostate tissues [10, 12, 30]. While such low expression of the ARVs challenges their physiological significance, a recent study of prostate cancer bone metastases shows that mRNA levels may not accurately reflect protein levels [18]. We did not assess protein levels of the ARVs

identified in the breast cancer cell lines because our primary aim was to examine the frequency and diversity of variant transcript expression. Future studies will build on the findings herein to determine which ARVs are commonly translated in the normal and malignant breast and other tissues, and how these factors impact on AR signalling in disease states.

It has been proposed that alternative splice variants of genes are rapidly degraded under physiological conditions but may be highly expressed under pathological conditions or other conditions of environmental stress if they are able to provide a selective advantage over their wild type counterparts [36]. This appears to be the case for truncated ARVs in prostate cancer. Androgen deprivation therapy for prostate cancer aims to inhibit classical ligand-mediated AR signalling, which requires an intact ligand binding domain. This treatment can select for ARVs that lack the LBD and are constitutively active, thus sustaining or resurrecting AR signalling in an androgen deplete environment or in the presence of anti-androgens [35, 37]. Based on evidence from preclinical studies that the AR can drive growth of a molecular subtype of breast cancer called molecular apocrine [38] or luminal androgen receptor (LAR) breast cancers [39], various forms of ADT (bicalutamide, enzalutamide, abiraterone) are currently in phase I/II clinical trials to examine the efficacy of treating women with ER-negative/AR-positive metastatic breast cancers with agents that repress classical ligand-mediated AR-signalling (see clinicaltrials.gov: NCT00468715, NCT01597193, and NCT00755885). In this study, we detected transcripts of multiple ARVs, including AR-V7, one of the most important ARVs in models of CRPC, in normal breast tissues and breast cancer cell lines, including two molecular apocrine/LAR breast cancer cell line models (MDA-453-MB and MFM223). Thus, the possibility of selection for or up-regulation of constitutively-active ARVs in breast tumours in response to ADT warrants further investigation.

In addition to prostate and breast cancers, AR also plays a role in other androgen-sensitive cancers including bladder, kidney, lung, and liver [21, 22]. For example, hepatocellular carcinoma (HCC) affects males more frequently than females and is considered to be androgen-dependent [40]. Hence, clinical trials examining the tumour suppressive efficacy of treatment with antiandrogens (e.g. flutamide) for HCC have been conducted [21, 41, 42]. These trials revealed no survival benefit for patients with advanced HCC. As we detected transcripts of multiple ARVs in the hepatocellular carcinoma HepG2 cell line, and liver tissue had the highest copies of AR-V7 transcripts, we speculate that using flutamide to treat men with HCC could select for ARV expression and concomitant androgen-independent AR-signalling. Our data provides credence for this concept, but the hypothesis remains to be investigated.

As discussed earlier, many ARVs (V1, V4, V5, V6, V7, and V9) share the NTD/DBD domains encoded by canonical exons 1-3 [10, 35]. The C-terminal short peptide of 1-53 amino acids encoded by the cryptic exons is the only structural difference that would contribute to their reported varying capacity for transcriptional activity [17, 30] and differing cellular localization [9, 10, 12, 30]. Thus, polymorphisms that change this peptide sequence would be expected to have potential impact on ARV biological function. In this study we have identified for the first time novel SNPs in four cryptic exons (CE1, CE3, CE4, and CE5) that could potentially lead to single amino acid substitutions at the C-terminal peptides of multiple previously reported ARVs. Chan and colleagues recently showed that K629A and R631A mutations shifted AR-V7 expression from predominantly nuclear to a mixed nuclear/cytoplasmic pattern [43]. The rs370756473G>T SNP leads to an R631L substitution, which is similar to the reported R631A mutation i.e. substitution of a neutral amino acid with a basic amino acid. Therefore, this R631L substitution might have similar effects as R631A mutation. Of note, the novel SNP (nt66831251A>G substitution) discovered in the present study leads to a V632I substitution at the C-terminally adjacent residue and thus might also modulate the activity of the encoded AR-V7 variant protein.

Previous studies have shown that CE2 is spliced to exon 3 using two different cryptic 3′ splice sites, which are only 80 bp apart in the *AR* gene, to give rise to AR-V5 and -V6 [11]. As shown in Fig. 4c, our results combined with the findings of Hu et al indicate that exon 9 uses four cryptic 3′ splice sites, which are only 135 bp apart in the *AR* gene, to generate six ARVs [10]. Classic introns have GT at the 5′ donor splice site, AG at the 3′ acceptor splice site, a splice branch site with a core consensus sequence of "CU(A/G)A(C/U)", and a polypyrimidine tract between the 3′ splice site and the branch site [44]. Among the four 3′ splice sites used by exon 9, two of them ($A^{66865241}$ and $G^{66865370}$) conform to the AG-rule and have a conserved branch site and a polypyrimidine tract. It appears that the 3′ splice site at $A^{66865241}$ is used more often than others as, through this site, exon 9 is spliced to upstream exons to give rise to four ARVs (V13, V14, V16, and V17). Neither of the $G^{66865369}$ and $G^{66865234}$ sites follows the AG-rule. Of interest, the site at $G^{66865370}$ used by AR-V15 is only 1 bp downstream of the previously reported site at $G^{66865369}$ used by AR-V12 [10]. We observed 4 clones carrying the AR-V15-specific splicing junction. Based on these observations, exon 9 is technically not a single exon but a group of "exons" with distinct 5′ coding sequences. Due to lack of antibodies recognizing the proteins encoded by the novel exon 9-containing transcripts, we were unable to examine the expression of these novel ARVs at protein level. Furthermore, the biological function of these ARVs remains to be determined in future studies.

In conclusion, we present evidence in this study that transcription of the *AR* gene gives rise to multiple alternatively spliced transcripts in breast cancer cell lines and normal human tissues. While the clinical relevance of these variant transcripts in breast cancer and other tissues awaits confirmation of protein expression, at least one of these splice variants (AR-V7) has documented capacity to encode a functional AR variant protein with clinical relevance in prostate cancer. Our findings suggest that further investigation of these ARVs and their clinical relevance in breast cancer and other androgen-sensitive diseases is warranted. In addition to revealing a broader scope of ARV transcript expression in human tissues, our study has identified novel ARV transcripts and previously unidentified single nucleotide substitutions in cryptic exons within the *AR* gene that could represent functional SNPs. These data provide a springboard to further study into how ARVs may function in a range of human pathologies.

Conflict of Interest The authors have no conflicts of interest to disclose.

References

1. Matsumoto T et al (2013) The androgen receptor in health and disease. Annu Rev Physiol 75:201–224
2. Chang C et al (2013) Androgen Receptor (AR) physiological roles in male and female reproductive systems: lessons learned from AR-knockout mice lacking AR in selective cells. Biol Reprod 89(1):21
3. Heinlein CA, Chang C (2004) Androgen receptor in prostate cancer. Endocr Rev 25(2):276–308
4. Fu W et al (2012) Progress of molecular targeted therapies for prostate cancers. Biochim Biophys Acta 1825(2):140–152
5. Scher HI et al (2004) Targeting the androgen receptor: improving outcomes for castration-resistant prostate cancer. Endocr Relat Cancer 11(3):459–476
6. Brand LJ, Dehm SM (2013) Androgen receptor gene rearrangements: new perspectives on prostate cancer progression. Curr Drug Targets 14(4):441–449
7. Gelmann EP (2002) Molecular biology of the androgen receptor. J Clin Oncol 20(13):3001–3015

8. Tilley WD et al (1989) Characterization and expression of a cDNA encoding the human androgen receptor. Proc Natl Acad Sci U S A 86(1):327–331

9. Dehm SM et al (2008) Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. Cancer Res 68(13):5469–5477

10. Hu R, Isaacs WB, Luo J (2011) A snapshot of the expression signature of androgen receptor splicing variants and their distinctive transcriptional activities. Prostate 71(15):1656–1667

11. Hu R et al (2009) Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. Cancer Res 69(1):16–22

12. Guo Z et al (2009) A novel androgen receptor splice variant is up-regulated during prostate cancer progression and promotes androgen depletion-resistant growth. Cancer Res 69(6):2305–2313

13. Ahrens-Fath I et al (2005) Androgen receptor function is modulated by the tissue-specific AR45 variant. FEBS J 272(1):74–84

14. Liu LL et al (2013) Mechanisms of the androgen receptor splicing in prostate cancer cells. Oncogene. doi:10.1038/onc.2013.284

15. Li Y et al (2012) AR intragenic deletions linked to androgen receptor splice variant expression and activity in models of prostate cancer progression. Oncogene 31(45):4759–4767

16. Nyquist MD et al (2013) TALEN-engineered AR gene rearrangements reveal endocrine uncoupling of androgen receptor in prostate cancer. Proc Natl Acad Sci U S A 110(43):17492–17497

17. Sun S et al (2010) Castration resistance in human prostate cancer is conferred by a frequently occurring androgen receptor splice variant. J Clin Invest 120(8):2715–2730

18. Hornberg E et al (2011) Expression of androgen receptor splice variants in prostate cancer bone metastases is associated with castration-resistance and short survival. PLoS One 6(4):e19059

19. Li Y et al (2011) Intragenic rearrangement and altered RNA splicing of the androgen receptor in a cell-based model of prostate cancer progression. Cancer Res 71(6):2108–2117

20. Liu G et al (2013) AR variant AR(v567es) induces carcinogenesis in a novel transgenic mouse model of prostate cancer. Neoplasia 15(9): 1009–1017

21. Chang C et al (2013) Androgen receptor (AR) differential roles in hormone-related tumors including prostate, bladder, kidney, lung, breast and liver. Oncogene. doi:10.1038/onc.2013.274

22. Siddique HR et al (2012) Androgen receptor in human health: a potential therapeutic target. Curr Drug Targets 13(14):1907–1916

23. Hickey TE et al (2012) Minireview: the androgen receptor in breast tissues: growth inhibitor, tumor suppressor, oncogene? Mol Endocrinol 26(8):1252–1267

24. Laurentino SS et al (2013) Identification of androgen receptor variants in testis from humans and other vertebrates. Andrologia 45(3):187–194

25. Zhu X et al (1997) Identification of an exon 3 deletion splice variant androgen receptor mRNA in human breast cancer. Int J Cancer 72(4): 574–580

26. Bhasker CR et al (2000) Genetic polymorphism of UDP-glucuronosyltransferase 2B7 (UGT2B7) at amino acid 268: ethnic diversity of alleles and potential clinical significance. Pharmacogenetics 10(8):679–685

27. Bhasker CR et al (1997) Allelic and functional variability of cytochrome P4502C9. Pharmacogenetics 7(1):51–58

28. McKinnon RA et al (1991) Localization of CYP1A1 and CYP1A2 messenger RNA in normal human liver and in hepatocellular carcinoma by in situ hybridization. Hepatology 14(5):848–856

29. Hu DG, Mackenzie PI (2009) Estrogen receptor alpha, fos-related antigen-2, and c-Jun coordinately regulate human UDP glucuronosyltransferase 2B15 and 2B17 expression in response to 17beta-estradiol in MCF-7 cells. Mol Pharmacol 76(2):425–439

30. Watson PA et al (2010) Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor. Proc Natl Acad Sci U S A 107(39):16759–16765

31. Gottlieb B et al (2012) The androgen receptor gene mutations database: 2012 update. Hum Mutat 33(5):887–894

32. Lim J, Ghadessy FJ, Yong EL (1997) A novel splice site mutation in the androgen receptor gene results in exon skipping and a non-functional truncated protein. Mol Cell Endocrinol 131(2):205–210

33. Hellwinkel OJ et al (1999) Complete androgen insensitivity caused by a splice donor site mutation in intron 2 of the human androgen receptor gene resulting in an exon 2-lacking transcript with premature stop-codon and reduced expression. J Steroid Biochem Mol Biol 68(1–2):1–9

34. Ris-Stalpers C et al (1994) Differential splicing of human androgen receptor pre-mRNA in X-linked Reifenstein syndrome, because of a deletion involving a putative branch site. Am J Hum Genet 54(4): 609–617

35. Dehm SM, Tindall DJ (2011) Alternatively spliced androgen receptor variants. Endocr Relat Cancer 18(5):R183–R196

36. Tress ML et al (2007) The implications of alternative splicing in the ENCODE protein complement. Proc Natl Acad Sci U S A 104(13): 5495–5500

37. Guo Z, Qiu Y (2011) A new trick of an old molecule: androgen receptor splice variants taking the stage?! Int J Biol Sci 7(6):815–822

38. Farmer P et al (2005) Identification of molecular apocrine breast tumours by microarray analysis. Oncogene 24(29):4660–4671

39. Doane AS et al (2006) An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. Oncogene 25(28):3994–4008

40. El-Serag HB, Rudolph KL (2007) Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology 132(7): 2557–2576

41. Chao Y et al (1996) Phase II study of flutamide in the treatment of hepatocellular carcinoma. Cancer 77(4):635–639

42. (2004) Randomized trial of leuprorelin and flutamide in male patients with hepatocellular carcinoma treated with tamoxifen. Hepatology 40(6):1361-1369

43. Chan SC, Li Y, Dehm SM (2012) Androgen receptor splice variants activate androgen receptor target genes and support aberrant prostate cancer cell growth independent of canonical androgen receptor nuclear localization signal. J Biol Chem 287(23):19736–19749

44. Zhuang YA, Goldstein AM, Weiner AM (1989) UACUAAC is the preferred branch site for mammalian mRNA splicing. Proc Natl Acad Sci U S A 86(8):2752–2756