Washington University School of Medicine
Digital Commons@Becker

2020-Current year OA Pubs | Open Access Publications

10-4-2022

Pan-cancer analysis reveals recurrent BCAR4 gene fusions across solid tumors

Andrew Nickless

Jin Zhang

Ghofran Othoum

Jace Webster

Matthew J Inkman

See next page for additional authors

Follow this and additional works at: https://digitalcommons.wustl.edu/oa_4

 Part of the Medicine and Health Sciences Commons

Authors

Andrew Nickless, Jin Zhang, Ghofran Othoum, Jace Webster, Matthew J Inkman, Emily Coonrod, Sherron Fontes, Emily B Rozycki, Christopher A Maher, and Nicole M White

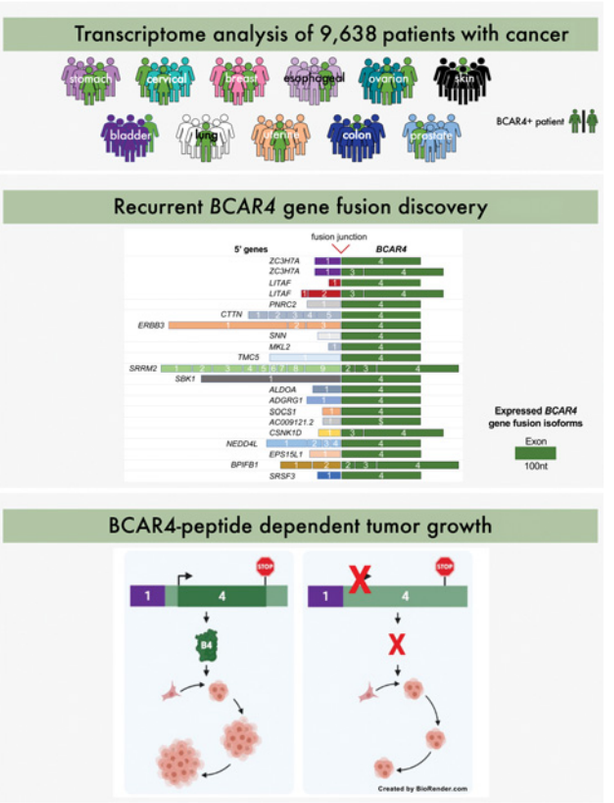
Pan-Cancer Analysis Reveals Recurrent *BCAR4* Gene Fusions across Solid Tumors

Andrew Nickless[1], Jin Zhang[2,3,4], Ghofran Othoum[1], Jace Webster[1], Matthew J. Inkman[2], Emily Coonrod[1], Sherron Fontes[1], Emily B. Rozycki[1], Christopher A. Maher[1,3,5], and Nicole M. White[1,3]

ABSTRACT

Chromosomal rearrangements often result in active regulatory regions juxtaposed upstream of an oncogene to generate an expressed gene fusion. Repeated activation of a common downstream partner–with differing upstream regions across a patient cohort–suggests a conserved oncogenic role. Analysis of 9,638 patients across 32 solid tumor types revealed an annotated long noncoding RNA (lncRNA), *Breast Cancer Anti-Estrogen Resistance 4 (BCAR4)*, was the most prevalent, uncharacterized, downstream gene fusion partner occurring in 11 cancers. Its oncogenic role was confirmed using multiple cell lines with endogenous *BCAR4* gene fusions. Furthermore, overexpressing clinically prevalent *BCAR4* gene fusions in untransformed cell lines was sufficient to induce an oncogenic phenotype. We show that the minimum common region to all gene fusions harbors an open reading frame that is necessary to drive proliferation.

Implications: *BCAR4* gene fusions represent an underappreciated class of gene fusions that may have biological and clinical implications across solid tumors.



Introduction

Chromosomal rearrangements are common somatic aberrations in cancer genomes, often leading to the juxtaposition of two genes, creating gene fusions. In addition to their biological roles in oncogenesis, some gene fusions are clinically relevant diagnostic markers, prognostic indicators, and therapeutic targets (1). Advances in next-generation sequencing coupled with improved gene fusion detection tools have accelerated the discovery of novel gene fusions in solid tumors. E26 transformation-specific (ETS) transcription factor family members (2), ALK and ROS fusions in lung cancer (3), and RAF kinase gene fusions in melanoma, gastric, and prostate cancer (4) represent important gene fusion discoveries across cancers. These exemplify a biological pattern called "functional recurrence": various 5′ regulatory regions join a particular 3′ oncogene across samples and cancer types. While a single gene fusion event may be infrequent and overlooked, functionally recurrent fusions activating the same downstream partner can accentuate highly relevant oncogenes and provide molecular insight into their function. Some gene fusions are low frequency

[1]Division of Oncology, Department of Medicine, Siteman Cancer Center, Washington University School of Medicine, St. Louis, Missouri. [2]Department of Radiation Oncology, Washington University School of Medicine, St. Louis, Missouri. [3]Siteman Cancer Center, Washington University School of Medicine, St. Louis, Missouri. [4]Institute for Informatics, Washington University School of Medicine, St. Louis, Missouri. [5]Department of Biomedical Engineering, Washington University School of Medicine, St. Louis, Missouri.

A. Nickless and J. Zhang are the co-first authors of this article.

C.A. Maher and N.M. White are the co-senior authors of this article.

Corresponding Authors: Christopher Maher, Washington University 660 S. Euclid Ave, Campus Box 8056, St. Louis, MO 63110. Fax: 314-362-9333; E-mail: christophermaher@wustl.edu; and Nicole White, nmmaher@wustl.edu

Mol Cancer Res 2022;20:1481–8

doi: 10.1158/1541-7786.MCR-21-0775

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

©2022 The Authors; Published by the American Association for Cancer Research

Downloaded from http://aacrjournals.org/mcr/article-pdf/20/10/1481/3210514/1481.pdf by Washington University St Louis user on 08 January 2023

(<5%) events in a single cancer (even with functionally recurrent events), but accumulation across solid tumors expands the potential biological and clinical relevance to a significantly broader patient population.

In this study, we applied INTEGRATE (5), our highly sensitive gene fusion discovery tool, in a pan-cancer strategy to detect functionally recurrent gene fusions. From our analysis of 9,638 patients across 32 different cancer types within The Cancer Genome Atlas (TCGA) consortium, we prioritized a novel class of *Breast Cancer Anti-Estrogen Resistance 4 (BCAR4)* gene fusions because: (i) they are the most prevalent gene fusions not previously characterized across solid tumors; (ii) full-length *BCAR4* is implicated in cancer (6, 7); and (iii) all patients express a minimum common region of *BCAR4* that contains an open reading frame (ORF), suggesting a conserved oncogenic role. Silencing the two most common fusions (*LITAF-BCAR4* and *ZC3H7A-BCAR4*) decreased proliferation in cancer cell lines; conversely, overexpressing the fusions in benign models increased it. Mutating the ORF abated *BCAR4* gene fusion-driven proliferation. Collectively, our pan-cancer analysis discovered a functionally recurrent class of *BCAR4* gene fusions that act through a protein to alter cell-cycle and proliferation across solid tumors.

Materials and Methods

TCGA datasets

We downloaded the aligned BAM files at https://portal.gdc.cancer.gov/ for the cancer types included in the TCGA Pan-Cancer Analysis listed in Supplementary Table S1. When available, RNA sequencing (RNA-seq) BAM files for matched adjacent normal tissue were also downloaded. Following TCGA practices, COAD and READ were merged to form a colorectal cancer cohort.

Gene fusion discovery

INTEGRATE is an open-source gene fusion discovery tool designed to map gene fusions using aligned RNA-seq reads and whole genome sequencing (WGS) paired-end sequencing reads, if available (5). INTEGRATE version 0.2.6 was run using default parameters in "RNA only" mode on the aligned RNA-seq reads (8). Analysis was based on hg38. Functionally recurrent 3′ fusion partners were identified if a gene was only found as a 3′ partner in somatic gene fusions across patients. Summary figures of novel and previously reported functionally recurrent 3′ fusion partners were generated using R (version 4.0.5).

Cell culture

SNU308 and TUHR14TKB were selected for their endogenous expression of *BCAR4* gene fusions with no expression of full length *BCAR4*. The TUHR14TKB cell line was purchased from RIKEN BRC Cell Engineering Division (RCB1383; RRID_CVCL_5953). SNU308 gallbladder cell line was purchased from the Korean Cell Line Bank (RRID:CVCL_5048). Cell lines were grown in RPMI-1640 (Genesee Scientific) supplemented with 10% FBS (Sigma) and 1% penicillin/streptomycin (Thermo Fisher Scientific). hTER1-HME1 were purchased from ATCC (ATCC, catalog no, CRL-10317, RRID: CVCL_0598). MCF10a were a gift from Dr. Ron Bose (Washington University; St. Louis, MO). Cell lines were cultured in MEGM Mammary Epithelial Cell Growth Medium BulletKit (Lonza) or following this protocol: https://brugge.med.harvard.edu/protocols. Cell line authentication and validation were performed by ATCC, RIKEN, and Korean Cell Line Bank. Cells were passaged less than 10 times and monitored for *Mycoplasma* by PCR.

Transfections and overexpression models

Cells were transfected with 10 nmol/L of custom silencer select siRNA (Supplementary Table S2) and Lipofectamine RNAiMax (Thermo Fisher Scientific) following the manufacturer's protocol and used for assays 72 hours later. Fusion constructs were synthesized by Invitrogen and Infusion cloned (Takara) into the pCFG5-IEGZ plasmid (a gift from Dr. Ron Bose). A terminal FLAG-tag was added to the *BCAR4* ORF and Infusion cloned. Sequences were confirmed by Sanger sequencing at Genewiz (South Plainfield). 293T cells were transfected with 3.75 μg of expression plasmid and an 8:1 ratio of pUMVC: VSVG. The next day media was exchanged for cell-specific complete media and virus collected at 48 and 72 hours. 2 mL of virus with 8 μg/mL Polybrene (Sigma) was added to mammalian cells, centrifuged at 2,500 RPM for 75 minutes, and fresh media exchanged after 6 hours. Cells were incubated with virus/polybrene for 6 hours the next day and used for assays 48 hours later. Point mutations within the ATG start site were introduced with Q5 site-directed mutagenesis Kit (NEB) to create the *LITAF-BCAR4* L-B mutant construct.

RNA isolation and cDNA synthesis

Total RNA was isolated with NucleoSpin RNA plus with DNA removal column (Macherey-Nagel). cDNA was synthesized with High Capacity cDNA Reverse Transcription Kit (Invitrogen) or 1-Step TB Green PrimeScript qRT-PCR kit (Takara).

qRT-PCR

Gene expression was confirmed with qRT-PCR using PowerSyBr Green (Invitrogen) or 1-Step TB Green PrimeScript (Takara). The comparative CT (ΔΔCT) method was used with values normalized to the housekeeping gene, *RPL32*, and to control samples. All primers (Supplementary Table S2) were obtained from Integrated DNA Technologies and determined to have 90% to 110% primer efficiency.

FITC Annexin V apoptosis detection

Cells were seeded at 200,000 cells/well and Annexin V staining determined 2 days later according to the manufacturer's protocol (BD Pharmingen).

Cellular nuclear and cytoplasmic isolation

Cells were isolated according to the PARIS kit protocol (Thermo Fisher Scientific) and gene expression determined by qPCR. Nuclear and cytoplasmic isolations were calculated by normalizing respective genes to total RNA expression.

Cell counting

Cells were seeded at 100,000 cells/well and counted every 2 days using the Countess II FL Automated Cell Counter (Thermo Fisher Scientific).

EdU proliferation assay

Cells were seeded at 250,000 cells/well and treated 2 days later with 10 μmol/L EdU (Thermo Fisher Scientific) for 2 hours. Cells were processed following the manufacturer's instructions and stained for DNA content with FxCycle Violet (Thermo Fisher Scientific). Analysis was performed on a FACScan flow cytometer (Becton Dickinson) at the Siteman Cancer Center Flow Cytometry Core (St. Louis, MO). A minimum of 10,000 events per sample were collected. FlowJo Version10 (RRID:SCR_008520; Becton Dickinson) was used to analyze data.

Downloaded from http://aacrjournals.org/mcr/article-pdf/20/10/1481/3210514/1481.pdf by Washington University St Louis user on 08 January 2023

Western blotting

Antibodies against FLAG (2368S; RRID:AB_2217020, 1:1000; Cell Signaling Technology), actin (3700S; RRID:AB_2242334, 1:10,000; Cell Signaling Technology), Goat anti-mouse or anti-rabbit peroxidase-conjugated secondary antibodies (Thermo Fisher Scientific) were used for experiments. Protein samples were prepared with ice-cold RIPA lysis buffer (25 mmol/L Tris pH7.5, 1% NP40, 0.1% SDS, 150 mmol/L NaCl, 0.5% sodium deoxycholate, and 1X Halt Protease Inhibitor Cocktail; Thermo Fisher Scientific). Lysate was subjected to a short sonication and then clarified by centrifugation at maximum speed, 4°C for 10 minutes. Protein concentration was determined with the DC Protein Assay (BioRad). Samples were diluted in loading buffer, boiled, and equal protein concentrations (20–30 μg) loaded and resolved by SDS-polyacrylamide gel electrophoresis in 12% or 4% to 12% Bolt Bis-Tris precast gels (Invitrgoen). Gels were transferred at 60°C for 1 hour to a nitrocellulose membrane (BioRad). Proteins were detected with specific antibodies and visualized and quantified on the ChemDoc MP Imaging System (BioRad) using secondary antibodies and Clarity Western ECL Substrate (Thermo Fisher Scientific).

Small peptide prediction and validation

For the proteogenomic search, we used ORFs predicted in transcripts of lncRNAs annotated in LNCipedia (9) appended to canonical proteins from Uniprot (RRID:SCR_002380; ref. 10) to ensure that the proteogenomic database is not biased towards the discovery of the *BCAR4* ORF. We used an equal number of reversed decoys to utilize the target-decoy strategy setting with a threshold of 0.01 for the FDR. Raw mass spectrometry files in mzML format were downloaded from the Clinical Proteomic Tumor Analysis Consortium data resource (CPTAC; RRID:SCR_017135, https://cptac-data-portal.georgetown.edu/datasets; ref. 11). Sequences between start codons (AUG, CUG, UUG) and stop codons (UAG, UGA, UAA) in each of the forward translated frames, with a minimum length of 100 nucleotides, were used as putative ORFs in lncRNAs for the construction of the proteogenomic database. This minimum threshold was selected to maximize signal to noise ratio: encompassing the discovery of known small ORFs (HOXB-AS3/LOC100507537) without increasing false positives from shorter peptides. Supplementary Table S3 shows the variable and fixed modifications used for each protein labeling protocol. The MSFragger search engine (12) was used allowing semitryptic peptides, two missed cleavage sites, $^{12}C/^{13}C$ isotope errors, and a precursor-ion mass tolerance of 20 ppm. PeptideProphet (RRID:SCR_000274) and ProteinProphet (RRID:SCR_000286; https://github.com/Nesvilab/philosopher; refs. 13, 14) were used to process the search engine results and infer protein groups, respectively. To further validate the novel peptides from the proteogenomic search, we followed the global FDR filtering step by peptide-centric validation as implemented in PepQuery (http://www.pepquery.org; ref. 15) and verified that the identified peptides from the *BCAR4* ORF are indeed statistically significant (PepQuery $P \leq 0.01$). PSIPRED Protein Analysis Workbench (RRID:SCR_010246; ref. 16) was used to predict protein secondary structure and InterProScan (RRID:SCR_005829; ref. 17) to identify protein domains.

Data availability

The data generated in this study are publicly available at https://github.com/ChrisMaherLab.

Results

Pan-cancer analysis discovers recurrent *BCAR4* fusions

We used INTEGRATE to analyze RNA-seq data from 9,638 patients across 32 different cancer types as part of TCGA consortium. To identify functionally recurrent gene fusion candidates, we prioritized genes that were recurrent 3′ gene fusion partners despite having different 5′ partners. As shown in **Fig. 1A**, the two most prevalent

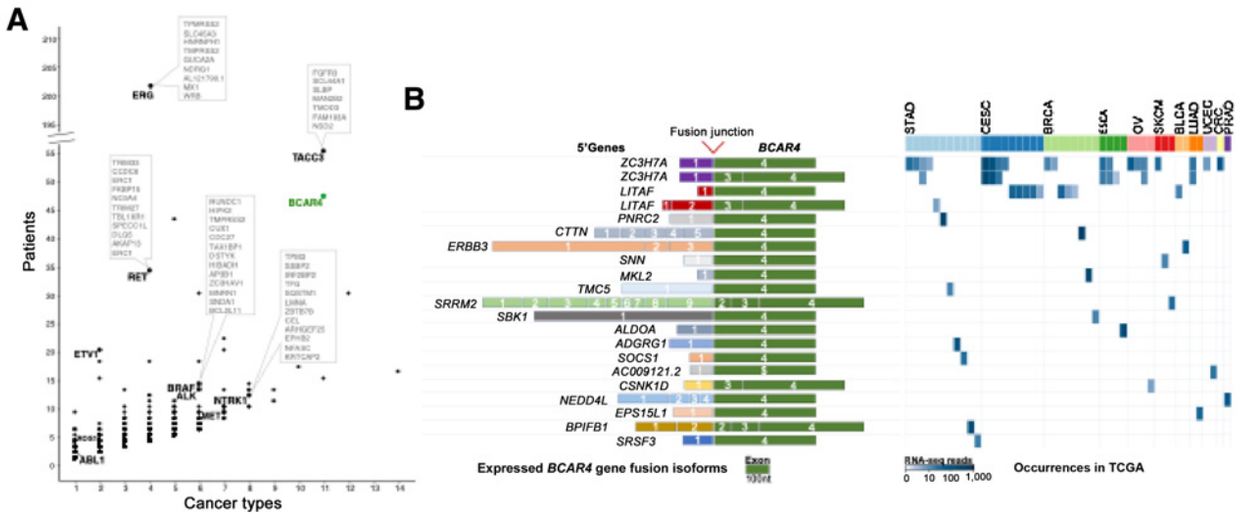


Figure 1.

Pan-cancer discovery of *BCAR4* gene fusions. **A,** Dot plot of recurrent gene fusions by number of patients across cancer types. Callout boxes list detected 5′ partners of known gene fusions. **B,** Left, Structure of expressed *BCAR4* fusion transcripts with the 5′ gene fusion partners represented in various colors and *BCAR4* represented in green. Fusions are sorted by descending prevalence. **B,** Right, Representation of patients across cancer types expressing *BCAR4* fusion transcripts. Rows correspond to the various *BCAR4* isoforms. A cell is colored blue if the patient expresses a particular *BCAR4* fusion, and the intensity corresponds to the number of RNA-seq reads supporting evidence of the fusion expression. STAD, stomach adenocarcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; BRCA, breast invasive carcinoma; ESCA, esophageal carcinoma; OV, ovarian serous cystadenocarcinoma; SKCM, skin cutaneous melanoma; BLCA, bladder urothelial carcinoma; LUAD, lung adenocarcinoma; UCEC, uterine corpus endometrial carcinoma; CRC, colorectal cancer; PRAD, prostate adenocarcinoma.

Downloaded from http://aacrjournals.org/mcr/article-pdf/20/10/1481/3210514/1481.pdf by Washington University St Louis user on 08 January 2023

recurrent 3′ gene fusion partners are ERG (202 patients) and TACC3 (55 patients). This is expected as ERG fusions with androgen-sensitive 5′ partners, including *TMPRSS2*, are highly recurrent gene fusions in prostate cancer. Further, *FGFR3-TACC3* is a highly recurrent gene fusion observed across multiple solid tumors. The third most prevalent class of gene fusions resulted in the expression of an annotated lncRNA–*BCAR4*–in 47 patients across 11 cancer types. *BCAR4* gene fusions are more prevalent than known clinically actionable recurrent gene fusions with immediate translational impact such as ABL1, BRAF, and ALK fusions.

The most common *BCAR4* gene fusion events are intrachromosomal rearrangements with breakpoints between the initial untranslated exons of either *ZC3H7A* or *LITAF* and the fourth exon of *BCAR4* (**Fig. 1B**). This results in the regulatory regions of *ZC3H7A* or *LITAF* minimally activating the expression of the fourth exon of *BCAR4*. Similarly, most expressed *BCAR4* gene fusion transcripts include early exons of 5′ partners spliced upstream of exon 4 of *BCAR4*. This functionally recurrent structure of *BCAR4* gene fusions across patients suggests exon 4 is the minimum region necessary to drive its function. Importantly, *BCAR4* gene fusions are tumor specific with no detection in adjacent normal tissues available from 718 TCGA patients.

Silencing *BCAR4* gene fusions decreases cell-cycle progression and proliferation in cancer cells

BCAR4 is a known oncogene promoting tumor growth (7), suggesting a cancer relevance for *BCAR4* gene fusions. To identify cancer cell lines harboring *BCAR4* fusions, we analyzed RNA-seq data from the Cancer Cell Line Encyclopedia (CCLE) using INTEGRATE and found two cell lines harboring the most common 5′ gene fusion partners: SNU308 gallbladder cancer cells (expressing *LITAF-BCAR4*) and TUHR14TKB renal carcinoma cells (expressing *ZC3H7A-BCAR4;* Supplementary Fig. S1A). These patient-derived lines, endogenously expressing only *BCAR4* fusions, are ideally suited to study the effects of *BCAR4* fusions on cancerous phenotypes.

To determine whether the *BCAR4* fusions drive cell-cycle progression in cancer cells, the fusions were transiently silenced with siRNAs and the cell-cycle profile assessed with flow cytometry to monitor EdU incorporation and DNA content. Knockdown of fusion expression resulted in significantly fewer S-phase cells and increase in G1 phase cells (**Fig. 2A** and **B**). Indeed, these fusion-targeting siRNAs consistently reduced the proportion of S-phase cells by 10% to 25% across these cell lines and increased the proportion of G1 phase cells by 5% to 19%. Knockdown of *BCAR4* gene fusions did not alter cell viability as determined by annexin staining (Supplementary Fig. S2A). These data show that endogenously expressed *BCAR4* gene fusions alter cell-cycle, specifically S-phase entry, in two different cancer cell lines.

BCAR4 gene fusions increase cell-cycle progression and proliferation in benign cells

Our data show that endogenous *BCAR4* fusions influence the behavior of cancer cells; next, we evaluated whether *BCAR4* gene fusions can drive proliferation in normal epithelial cell lines. We hypothesized that expression of full length *BCAR4* or the common *BCAR4* fusions would drive cell-cycle progression resulting in more proliferating (S-phase) cells and fewer G1-phase cells. Full length *BCAR4*, the *LITAF-BCAR4* (L-B fusion), or the *ZC3H7A-BCAR4* fusion (Z-B fusion) were introduced into untransformed cell lines (HME1 and MCF10a; Supplementary Fig. S1B) and expression levels were monitored by qRT-PCR (Supplementary Fig. S3); the cell-cycle profile and proliferation were then assessed. Cell-cycle analysis confirmed that HME1 cells expressing full length *BCAR4* had 29% more S-phase cells than empty vector (EV) controls (Supplementary Fig. S4A). Interestingly, the L-B and Z-B fusions more efficaciously drove HME1 cells into S-phase (56% and 65% increases, respectively; **Fig. 2C**). Similarly, expression of *BCAR4* gene fusions in MCF10a cells significantly increased the proportion of S-phase cells (greater than 80% relative to EV; **Fig. 2D**). There were no concurrent changes in cell viability (Supplementary Fig. S2B). These results demonstrate *BCAR4* fusion expression can promote cell-cycle progression in benign cell lines.

To determine whether the observed increase in cell-cycle progression leads to an increase in cellular proliferation, the growth of full length *BCAR4*- and fusion-overexpressing cells were monitored with cell counting. Significant increases in the number of HME1 cells expressing the L-B and Z-B fusions were observed starting at 48 hours, and at 96 hours there were approximately 3-fold more cells relative to EV controls (**Fig. 2C**, right). Expression of full length *BCAR4* yielded more modest results (<2-fold more than EV at 96 hours; Supplementary Fig. S4B). MCF10a *BCAR4*-fusion cells also had increased cellular proliferation (**Fig. 2D**, right). These data indicate that *BCAR4* gene fusion expression is sufficient to increase cell growth in two normal cell lines.

BCAR4 encodes a small peptide responsible for the oncogenic phenotypes

Since the fourth exon of *BCAR4* is the minimum common region found in all gene fusion events across cancers, we speculated that it was functionally important. Our sequence analysis of *BCAR4* identified a previously reported small peptide (18). The *BCAR4* ORF is 366 nucleotides in length and produces a 121 amino acid peptide chain predicted to have structure and multiple transmembrane domains (Supplementary Fig. S5). Evidence of the BCAR4 peptide in cancer patient samples was found by mining publicly available mass spectrometry data. Two different semi-tryptic digestion fragments corresponding to the BCAR4 protein were reliably detected in at least four cancer types (**Fig. 3A**; Supplementary Fig. S6A and S6B; Supplementary Table S4). In addition, ribosome profiling (Ribo-seq) analysis from public datasets (19) shows a focal enrichment in the predicted ORF within exon 4 of the *BCAR4* transcript (Supplementary Fig. S6C). These data are strong evidence that *BCAR4* encodes an expressed protein in solid tumors.

The BCAR4 protein contributes to cancer phenotypes (18); however, there is also evidence that *BCAR4* functions as a lncRNA (7). For insight into the function of *BCAR4*, we assayed the localization of *BCAR4* fusion transcripts: cytoplasmic localization would indicate translation potential while nuclear localization would indicate a noncoding function of *BCAR4*. Endogenously expressed *BCAR4* fusion transcripts localized predominantly to the cytoplasm: 89% of *LITAF-BCAR4* RNA in SNU308 cells and 62% of *ZC3H7A-BCAR4* RNA in TUHR14TKB cells (Supplementary Fig. S7). These data are consistent with translation of the *BCAR4* ORF.

To directly test the functional significance of the *BCAR4* ORF, we mutated the start codon in the *LITAF-BCAR4* fusion plasmid, thus abrogating potential translation of the *BCAR4* ORF (Supplementary Fig. S1B). HME1 and MCF10a cells expressing the mutant L-B fusion had significantly fewer S-phase cells than wild-type L-B fusion expressing cells, indicating that translation of the ORF is at least partly responsible for the cell-cycle phenotype (**Fig. 3B** and **C**). Over time these differences in cell-cycle progression led to large changes in cell number: 96 hours after plating, there were twice as many HME1 cells expressing the L-B fusion than the mutant L-B fusion (**Fig. 3B**, right). Changes in cell-cycle and cell number were also observed in MCF10a cells (**Fig. 3C**, right). To directly detect BCAR4 protein in our models,

Downloaded from http://aacrjournals.org/mcr/article-pdf/20/10/1481/3210514/1481.pdf by Washington University St Louis user on 08 January 2023

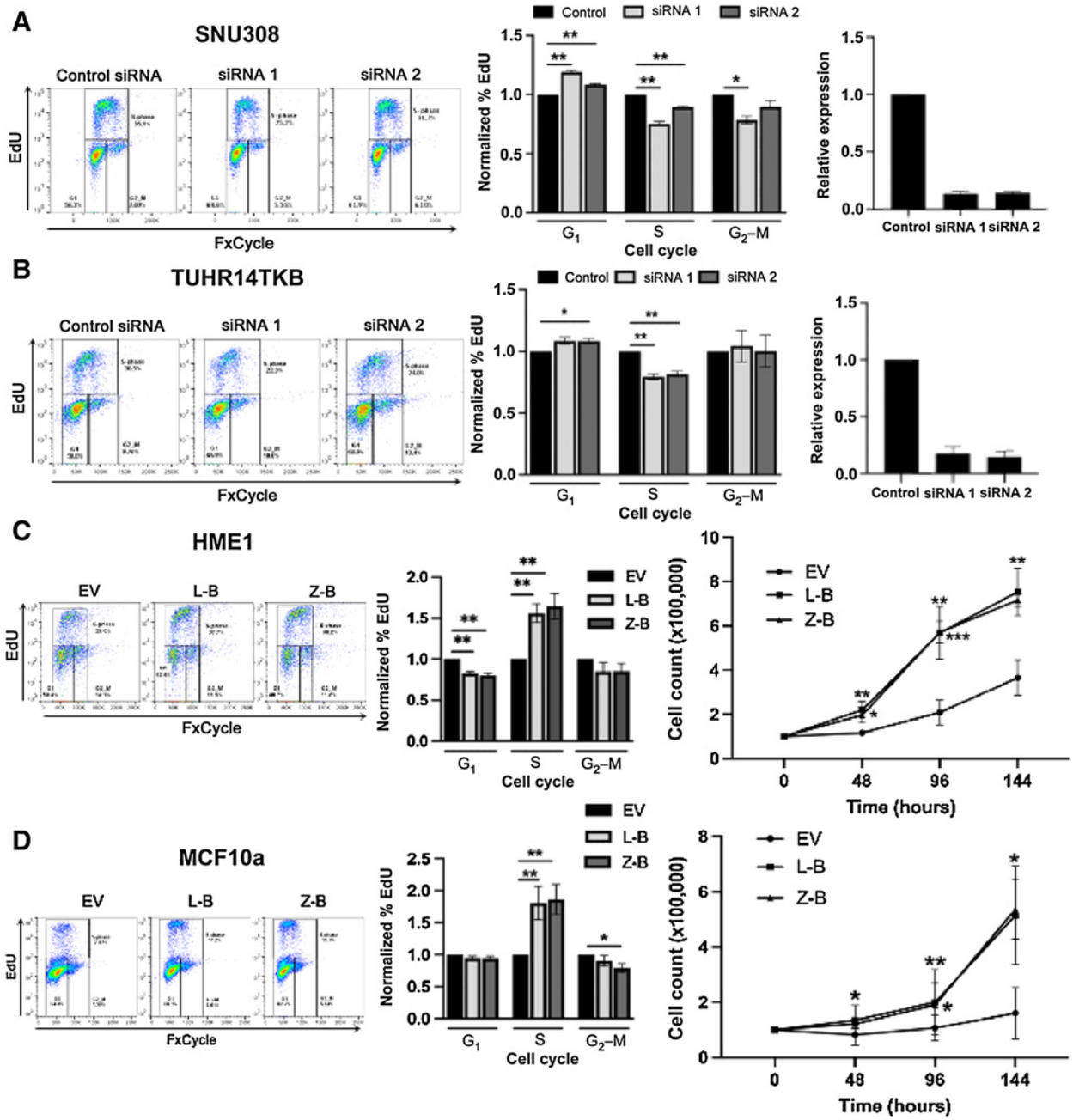


Figure 2.

BCAR4 gene fusions alter cell-cycle and proliferation. Representative dot plots of EdU-DNA stain (FxCycle) flow cytometry analysis and quantification of G1, S, G_2–M cell populations after siRNA-mediated silencing of *BCAR4* fusions in SNU308 ($n = 3$; **A**) or TUHR14TKB ($n = 4$; **B**) cells. **A** and **B,** Right, qRT-PCR analysis confirmed knockdown of *BCAR4* fusions. Representative dot plots of EdU-DNA stain (FxCycle) flow cytometry analysis and quantification of G1, S, G_2–M cell populations overexpressing EV or gene fusion (L-B or Z-B fusion) in HME1 ($n = 4$; **C**) or MCF10a cells ($n = 6$; **D**). Bar graphs present normalized mean ± SEM. Paired ratiometric two-tailed *t* tests were performed. **C** and **D,** Right, Cell growth curve analysis of HME1 ($n = 5$) or MCF10a ($n = 6$) cells overexpressing EV, L-B, or Z-B fusions. Graphs present mean ± SEM. Paired two-tailed *t* tests were performed. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

we FLAG-tagged the constructs and observed robust expression in cells expressing the L-B fusion but not the mutant L-B fusion (**Fig. 3D**; Supplementary Fig. S8). Some residual signal was observed from the mutant L-B construct, potentially caused by use of a second in-frame ATG codon just downstream of the canonical start codon. Expression of the *BCAR4* ORF alone (without additional noncoding sequence) similarly promoted cell-cycle progression and proliferation in MCF10a cell models (Supplementary Fig. S9). Overall, these data show that the fourth exon of *BCAR4* can produce a small peptide capable of inducing cell growth.

Downloaded from http://aacrjournals.org/mcr/article-pdf/20/10/1481/3210514/1481.pdf by Washington University St Louis user on 08 January 2023

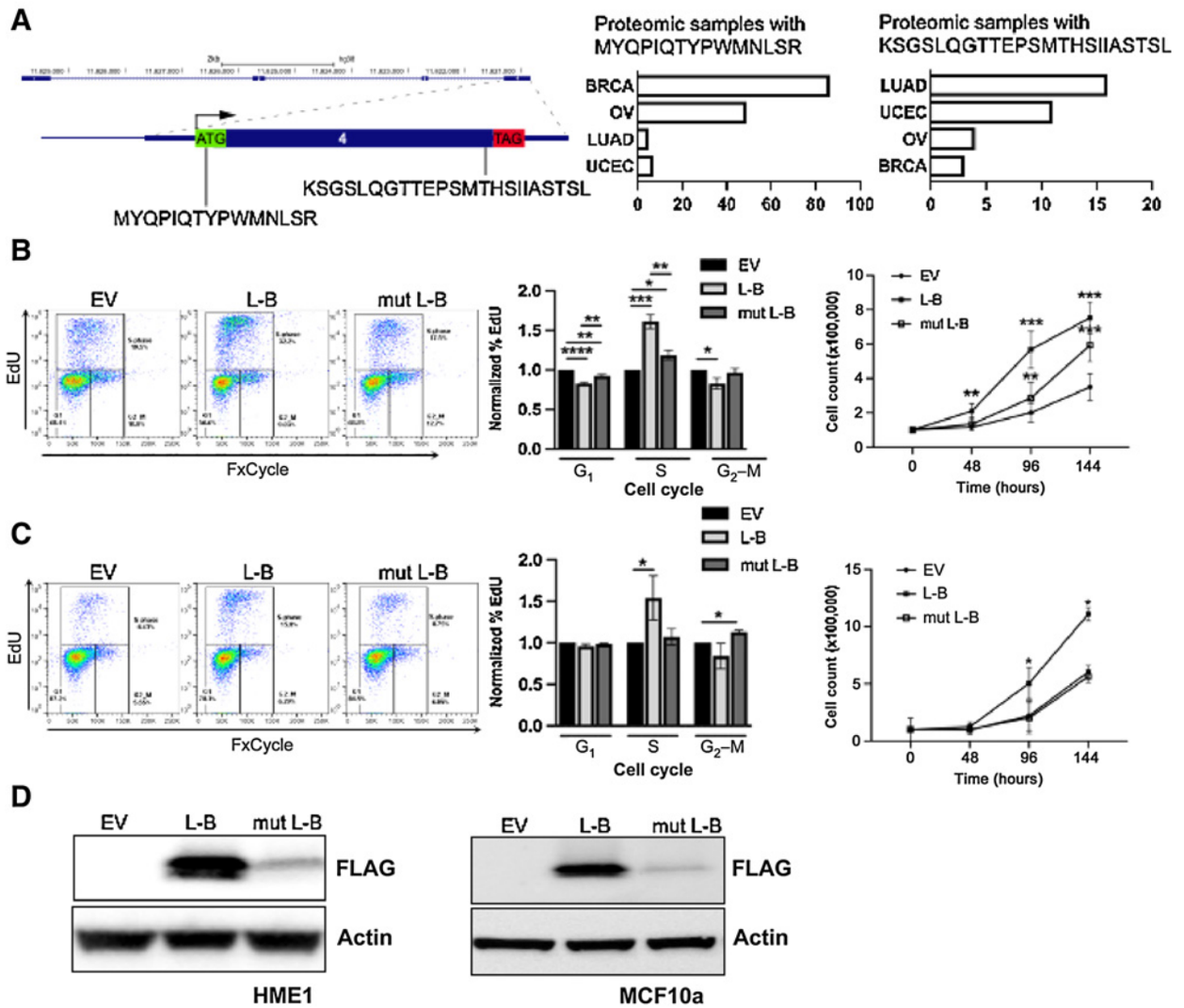


Figure 3.

BCAR4 encodes a small protein that alters cell-cycle and proliferation. **A,** Left, Schematics depicting the *BCAR4* gene, its predicted open reading frame in exon 4, and the tryptic peptides of *BCAR4* detected by mass spectrometry. **A,** Right, BCAR4 peptide support in lung adenocarcinoma (LUAD), uterine corpus endometrial carcinoma (UCEC), ovarian serous cystadenocarcinoma (OV), and breast invasive carcinoma (BRCA). Representative dot plots of EdU-DNA stain (FxCycle) flow cytometry analysis and quantification of G1, S, G_2–M cell populations of **B,** HME1 ($n = 6$) and **C,** MCF10a ($n = 3$) overexpressing EV, L-B, or mutant L-B (mut L-B) fusions. Bar graphs present normalized mean ±SEM. Paired ratiometric two-tailed *t* tests were performed. **B** and **C,** Right, Cell growth curve analysis of HME1 ($n = 6$) or MCF10a cells ($n = 4$) overexpressing EV, LB, or mut L-B. Graphs present mean ±SEM. Paired two-tailed *t* tests were performed. **D,** Western blot of FLAG-tagged *BCAR4* ORF expression in HME1 and MCF10a *BCAR4*-fusion–overexpressing cells, $n = 4$. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

Discussion

Using our highly sensitive gene fusion discovery tool, INTEGRATE, we discovered a novel gene fusion activating a known oncogene, *BCAR4*. This is the first unbiased pan-cancer study to systematically discover functionally recurrent *BCAR4* gene fusions across cancer types. These findings have clinical significance as (i) full-length *BCAR4* is implicated in poor clinical outcome and treatment resistance (6); (ii) full-length *BCAR4* has oncogenic potential in breast cancer (7, 18) and NSCLC (20); and (iii) focal amplifications of *BCAR4* were found in 20% of cervical patients in the TCGA dataset (21). Our findings extend the clinical significance of *BCAR4* into new cancer types–including highly aggressive cancers–to provide an important mechanism of how *BCAR4* induces oncogenic properties.

Our findings indicate that *BCAR4* chimeras are oncogenic driver events across cancer types. We discovered 47 *BCAR4* gene fusions with 19 different 5′ partners–mostly involving only the first untranslated regulatory exon of the 5′ gene. A previous study in NCSLC also found *BCAR4* gene fusions with the 5′ partner CD63 (22). Our pan-cancer study expands beyond lung cancer to discover *BCAR4* gene fusion events in 11 cancers, which surpasses the prevalence of well characterized functionally recurrent gene fusions such as EVT and ROS (two

Downloaded from http://aacrjournals.org/mcr/article-pdf/20/10/1481/3210514/1481.pdf by Washington University St Louis user on 08 January 2023

cancers), ERG (four cancers), BRAF and ALK (six cancers). The consistent expression of exon 4 of *BCAR4* suggests functional recurrence and implicates a biological and clinical importance of *BCAR4* gene fusions in solid tumors.

It is challenging to confidently annotate RNAs as coding or noncoding with growing evidence suggesting RNAs may encode small ORFs producing stable, functional peptides. Complicating this process is the possibility that RNAs can have both coding and noncoding functions (23–26). Full length *BCAR4* was first classified as a protein with expression in human oocytes and placenta (27); however, more recently is it described as a lncRNA (7), raising the possibility that *BCAR4* may function as a noncoding RNA and/or a protein-coding gene. Here, we show that the minimum common region of *BCAR4* across all gene fusions is the fourth exon, which harbors an ORF. Our computational and experimental data provide evidence for translation of the ORF to produce a functional protein capable of influencing oncogenic phenotypes. This is supported by detecting BCAR4 peptide in patient tumors and FLAG-tagged BCAR4 protein in cell culture models. Fusion and ORF-only constructs had significantly increased proliferation in benign cell lines while a construct with a mutated ATG start site had significantly diminished protein expression and proliferation. Genomic editing could provide additional evidence for the importance of the peptide relative to the lncRNA function. These results establish *BCAR4* fusions and their resulting peptide products as potential oncogenic drivers and possible therapeutic targets.

Authors' Disclosures

A. Nickless reports a patent for BCAR4 as a predictive biomarker and therapeutic target across cancers pending. J. Zhang reports a patent for BCAR4 as a predictive biomarker and therapeutic target across cancers pending. C.A. Maher reports grants from Emerson Collective; and grants from The Alvin J. Siteman Cancer Center Siteman Investment Program, The Foundation for Barnes-Jewish Hospital Cancer Frontier Fund, and Barnard Trust during the conduct of the study; other support from Tempus outside the submitted work; in addition, C.A. Maher has a patent for BCAR4 as a predictive biomarker and therapeutic target across cancers pending. N.M. White reports a patent for BCAR4 as a predictive biomarker and therapeutic target across cancers pending. No disclosures were reported by the other authors.

Authors' Contributions

A. Nickless: Conceptualization, data curation, formal analysis, validation, investigation, writing–original draft, writing–review and editing. **J. Zhang:** Software, formal analysis, visualization. **G. Othoum:** Formal analysis, investigation, visualization, writing–review and editing. **J. Webster:** Formal analysis, investigation, writing–review and editing. **M.J. Inkman:** Software, writing–review and editing. **E. Coonrod:** Investigation, methodology, writing–review and editing. **S. Fontes:** Investigation. **E.B. Rozycki:** Investigation. **C.A. Maher:** Conceptualization, formal analysis, supervision, funding acquisition, visualization, writing–review and editing. **N.M. White:** Conceptualization, formal analysis, supervision, validation, visualization, writing–original draft.

Acknowledgments

C.A. Maher received funding from The Alvin J. Siteman Cancer Center Siteman Investment Program through The Foundation for Barnes-Jewish Hospital Cancer Frontier Fund (grant no. GR0012692) and Barnard Trust (grant no. GR0008791), and The Emerson Collective (grant no. GR0017807). We thank the Alvin J. Siteman Cancer Center at Washington University School of Medicine and Barnes-Jewish Hospital (St. Louis, MO), for the use of the Siteman Flow Cytometry. The Siteman Cancer Center is supported in part by an NCI Cancer Center Support Grant (grant no. P30 CA091842).

Note

Supplementary data for this article are available at Molecular Cancer Research Online (http://mcr.aacrjournals.org/).

Received September 13, 2021; revised April 4, 2022; accepted June 10, 2022; published first June 16, 2022.

References

1. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun X-W, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005;310:644–8.
2. Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, et al. Transcriptome sequencing to detect gene fusions in cancer. Nature 2009;458:97–101.
3. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell 2012;150:1121–34.
4. Palanisamy N, Ateeq B, Kalyana-Sundaram S, Pflueger D, Ramnarayanan K, Shankar S, et al. Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. Nat Med 2010;16:793–8.
5. Zhang J, White NM, Schmidt HK, Fulton RS, Tomlinson C, Warren WC, et al. INTEGRATE: gene fusion discovery using whole genome and transcriptome data. Genome Res 2016;26:108–18.
6. Godinho MFE, Sieuwerts AM, Look MP, Meijer D, Foekens JA, Dorssers LCJ, et al. Relevance of BCAR4 in tamoxifen resistance and tumour aggressiveness of human breast cancer. Br J Cancer 2010;103:1284–91.
7. Xing Z, Lin A, Li C, Liang K, Wang S, Liu Y, et al. lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. Cell 2014;159:1110–25.
8. Lau JW, Lehnert E, Sethi A, Malhotra R, Kaushik G, Onder Z, et al. The cancer genomics cloud: collaborative, reproducible, and democratized-a new paradigm in large-scale computational research. Cancer Res 2017;77:e3–6.
9. Volders P-J, Helsens K, Wang X, Menten B, Martens L, Gevaert K, et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res 2013;41:D246–251.
10. UP Consortium. UniProt: a hub for protein information. Nucleic Acids Res 2015;43:D204–212.
11. Edwards NJ, Oberti M, Thangudu RR, Cai S, McGarvey PB, Jacob S, et al. The CPTAC data portal: a resource for cancer proteomics research. J Proteome Res 2015;14:2707–13.
12. Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. Nat Methods 2017;14:513–20.
13. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 2002;74:5383–92.
14. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 2003;75:4646–58.
15. Wen B, Wang X, Zhang B. PepQuery enables fast, accurate, and convenient proteomic validation of novel genomic alterations. Genome Res 2019;29:485–93.
16. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Res 2019;47:W402–7.
17. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. Nucleic Acids Res 2005;33:W116–120.
18. Godinho M, Meijer D, Setyono-Han B, Dorssers LCJ, van Agthoven T. Characterization of BCAR4, a novel oncogene causing endocrine resistance in human breast cancer cells. J Cell Physiol 2011;226:1741–9.
19. Michel AM, Fox G, M Kiran A, De Bo C, O'Connor PBF, Heaphy SM, et al. GWIPS-viz: development of a ribo-seq genome browser. Nucleic Acids Res 2014;42:D859–864.
20. Yang H, Yan L, Sun K, Sun X, Zhang X, Cai K, et al. lncRNA BCAR4 increases viability, invasion, and migration of non-small cell lung cancer cells by targeting glioma-associated oncogene 2 (GLI2). Oncol Res 2019;27:359–69.
21. Cancer Genome Atlas Research Network, Albert Einstein College of Medicine, Analytical Biological Services, Barretos Cancer Hospital, Baylor College of Medicine, Beckman Research Institute of City of Hope, et alet al. Integrated

Downloaded from http://aacrjournals.org/mcr/article-pdf/20/10/1481/3210514/1481.pdf by Washington University St Louis user on 08 January 2023

genomic and molecular characterization of cervical cancer. Nature 2017;543:378–84.
22. Bae K, Kim JH, Jung H, Kong S-Y, Kim YH, Kim S, et al. A fusion of CD63-BCAR4 identified in lung adenocarcinoma promotes tumorigenicity and metastasis. Br J Cancer 2021;124:290–8.
23. Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, et al. Peptidomic discovery of short open reading frame-encoded peptides in human cells. Nat Chem Biol 2013;9:59–64.
24. Chen J, Brunner A-D, Cogan JZ, Nuñez JK, Fields AP, Adamson B, et al. Pervasive functional translation of noncanonical human open reading frames. Science 2020;367:1140–6.
25. Othoum G, Coonrod E, Zhao S, Dang HX, Maher CA. Pan-cancer proteogenomic analysis reveals long and circular noncoding RNAs encoding peptides. NAR Cancer 2020;2:zcaa015.
26. Prensner JR, Enache OM, Luria V, Krug K, Clauser KR, Dempster JM, et al. Noncanonical open reading frames encode functional proteins essential for cancer cell survival. Nat Biotechnol 2021;39:697–704.
27. Angulo L, Perreau C, Lakhdari N, Uzbekov R, Papillier P, Freret S, et al. Breast-cancer anti-estrogen resistance 4 (BCAR4) encodes a novel maternal-effect protein in bovine and is expressed in the oocyte of humans and other non-rodent mammals. Hum Reprod Oxf Engl 2013;28:430–41.

Downloaded from http://aacrjournals.org/mcr/article-pdf/20/10/1481/3210514/1481.pdf by Washington University St Louis user on 08 January 2023