Supplemental data

Bioinformatic analyses of ORF of mRSK4 mRNAs

Since the NM_025949.2 mRSK4 is shown as a DNA sequence, just like all other mRNA sequences in the NCBI database, herein AUG start codon is converted to ATG (Fig. 1). The first start codon in the NM_025949.2 sequence resides at the 144-146th nt, dubbed as ATGα which initiates translation of an 860-aa protein that ends at a TGA stop codon at the 2724-2726th nt (S-Fig. 3 and S-Fig. 4). There are additional 20 ATGs downstream of the ATGα that are in the same reading frame initiated from ATGα (S-Fig. 3 and S-Fig. 4). The possible mRSK4 protein isoforms initiated from these downstream ATGs are truncated at the N-terminus, with their length listed in the left table of S-Fig. 4.

The my090 mouse mRNA we cloned lacks 199 nt of the 5'-end of the NM_025949.2 sequence and thus lacks the ATGα (Fig. 1). Its first ORF encodes a 15-aa peptide initiated from an ATG at the 178-180th nt and ending at a TGA stop codon at the 223-225th nt (ATG1 and TGA1 in my090 in S-Fig. 1). There is another ATG just 1 nt downstream of the ATG1, which is coined as ATGγ because we name the first ATG in the NM_014496.4 hRSK4 as ATGβ and this ATGβ also appears in both my090 (at the 182-184th nt) and NM_025949.2 (at the 381-383rd nt) sequences (Fig. 1, S-Fig. 3, and the left table in S-Fig. 4). ATGγ and ATGβ are in the same reading frame, but the protein initiated from ATGγ has additional 17 aa at the N-terminus (S-Fig. 2C). However, the ATGγ is actually within the 15-aa upstream ORF (uORF), and its -3 nt is a primidine (underlined T in Fig. 1) that makes the ATGγ less efficient, according to the principle of Kozak sequence.[32] On the other hand, the -3 nt of the ATGβ in both mice and humans are a purine (underlined G in Fig. 1) that makes a perfect Kozak sequence.[32] Therefore, it remains possible that the my090 sequence may actually be initiated from ATGβ via a leaky scanning and/or a reinitiation mechanism with the ATG1 and ATGγ as upstream ATGs (uATG). It also remains possible that ATGβ may be an additional authentic start codon in the NM_025949.2 sequence, besides the ATGα, simultaneously producing two proteins. Notably, in all three putative proteins initiated from ATGα, ATGγ or ATGβ, the second aa is leucine (encoded by CTG or CTA; Fig. 1), suggesting that the efficiency for removal of the initiator methionine is similar for all three proteins.

Bioinformatic analyses of ORF of hRSK4 mRNAs

The NM_014496.4 hRSK4 sequence lacks the ATGα and ATGγ (Fig. 1). ATGβ is its first start codon (at the 37-39th nt) that initiates a 745-aa protein (the right table in S-Fig. 4). However, the my092 variant we cloned has an additional 222 nt at the 5'-end of the NM_014496.4 sequence, which introduces in the uATG1 at the 144-146th nt and the uATG2 at the 201-203rd nt. These two uATGs are within the same reading frame that ends at the 249-251st nt (TGA1 in S-Fig. 3) and thus produces a peptide of 35 aa and 16 aa, respectively. The TGA1 is just 7 nt upstream of the ATGβ (S-Fig. 3). It is possible that the my092 variant is still translated from ATGβ but the translation efficiency is controlled by the 35-aa and/or the 16-aa uORFs via such as leaky scanning or reinitiation. Moreover, both my092 and NM_014496.4 sequences have 18 additional in-frame ATGs downstream of ATGβ (the right table in S-Fig. 4) which may initiate translation of proteins that lack part of the N-terminus.

In the variant of hRSK4 that contains exons 1B, 1C and 1D, exon 1B introduces in a TGA stop codon to terminate the translation from ATGβ (ATG1 in S-Fig. 3, panel 3), producing a 44-aa peptide. Actually, within this 44-aa ORF there is an in-frame ATG (ATG2 in S-Fig. 3, panel

3) that is 42 nt downstream of ATG1, initiating a 30-aa peptide. Moreover, exon 1D introduces in a second short uORF that has only 9 nt (three triplets) in length, starting from atg3 and ending at tag2 (S-Fig. 3, panel 3). Exon 1D also brings in a new ATG named ATGδ (Fig. 1, S-Fig. 2C and S-Fig. 3), which locates at only 19 nt downstream of the short uORF2 and also initiates translation of a 745-aa hRSK4. The first 27 aa of the 44-aa peptide encoded by uORF1 are derived from exon 1, while the coding region of exon 1D initiated from ATGδ encodes also 27 aa. As a result, the first 27 aa of the 745-aa hRSK4 protein initiated from ATGδ differs from the first 27 aa of the 745-aa hRSK4 initiated from the ATGβ in the NM_025949.2 sequence (S-Fig. 2C and S-Fig. 3). An uncertainty is whether translation can still be initiated from ATGδ, after termination of the ATGβ-initiated uORF1 of the 44-aa peptide and termination of the three-triplet ORF2, since Kozak showed that reinitiation could barely be detectable when an uORF is longer than 35 aa.[32]

In the variant that contains exon 1D but not exons 1B and 1C (my106 in Fig. 1), exon 1D encompasses the 9-nt (atgttttag) short uORF (S-Fig. 3, panel 4) as described above. Moreover, exon 1D also introduces in another taa stop codon (taa2 in S-Fig. 3, panel 4) that terminates the ORF initiated from the ATGβ (ATG1 in S-Fig. 3, panel 4) or from its downstream ATG2, producing a 35-aa or 21-aa peptide, respectively. However, unlike the situation in my111 sequence, taa2 is just 2 nt upstream of the ATGδ in exon 1D. This closer relationship may allow the ATGβ-initiated ORF to serve as a better uORF for reinitiation at ATGδ, compared with the situation in my111 (S-Fig. 3, panel 3). Like in my111, the ATGδ in my106 sequence also initiates a 745-aa hRSK4 protein, the first 27 aa of which differ from the first 27 aa of the protein initiated from ATGβ in the NM_025949.2 sequence (S-Fig. 2C). Like other above described variants, my111 and my106 have the same in-frame ORFs downstream of the ATGδ.

Theoretically, the 222 nt at the 5'-end of the my092 sequence may appear in the mRNAs with or without exon 1D (my106) or exons 1B, 1C and 1D (my111), since there is no indication suggesting that these mRNA variants are mutually exclusive. Variants simultaneously containing the 222 nt in the 5' end of exon 1 or exons 1B, 1C and 1D have more complex uORFs, as illustrated in an assembled my106 sequence (the bottom panel in S-Fig. 3).

Besides the above-described translation start codons, alternative splicing at the 15th nt of exon 22 in mouse or at the penultimate exon in both human and mouse can also lead to in-frame deletion and thus produce different protein isoforms (S-Fig. 2C). The deletion of 13 aa in the penult exon induced by indolocarbazole derivatives may further increase the number of hRSK4 protein isoforms in abnormal situations (S-Fig. 2C).

Supplementary Materials and Methods

Cell lines: 168FARN, 4T07, 4T1, 66C14 and 67RN are cell lines established from the same mouse mammary tumor.[65] NMuMG and HC11 are non-malignant mouse mammary epithelial cell lines. Ela-mycPT1 and Ela-mycPT4 are cell lines we established from two different Ela-*myc* transgenic pancreatic tumors.[25,26] M8 is a pcDNA3.1-c-*myc* expressing clone of the Ela-mycPT4 cells.[25,26] PT1-vec and PT1-D1 are a pcDNA3.1-neo empty vector clone and a pcDNA3.1-cyclin D1 (D1) clone, respectively, of Ela-mycPT1 cells. MCF7, MB231, T47D, and SKBR3 are human breast cancer cell lines. MCF10 is a non-malignant human breast epithelial cell line. MCF15 is a new human breast cancer cell line in the MCF series.[66] GI101 is a human breast cancer cell line provided by Dr. JE Price at M.D. Anderson Cancer Center, University of Texas. AsPC-1, Colo357, Panc-28 and L3.6pL are human pancreatic cancer cell lines. HL60 is a human promyelocytic leukemia cell line. The p53-/- single and the p53-mdm2 double null mouse

embryonic fibroblast (MEF) cell lines were provided by Dr. Guillermina Lozano at MD Anderson Cancer Centre. PTEN-/- and -/+ MEFs were provided by Dr. Tak Mak at Campbell Family Cancer Research Institute in Canada. P16-/-MEF was provided by Dr. Norman E. Sharpless at the University of North Carolina School of Medicine. CCND1-/-MEF was provided by Dr. Alan Diehl at University Of Pennsylvania. RSK2-/-MEF was provided by Dr. Jens Brüning at University of Cologne in Germany. All these MEFs were used with the providers' permission. There were many more mouse and human cell lines that were used in this study but were not mentioned, because the results were not particularly different from those with the above described cell lines and thus were not presented.

Sources of tissue, RNA and cDNA library: MMTV-c-*myc* mammary tumor tissues and Ela-c-*myc* pancreatic tumor tissues were collected from the corresponding transgenic mice as reported before.[67,68] Normal tissues of various mouse organs were corrected from the adult non-transgenic littermates. Total RNA samples of various normal human organs were purchased from Clontech (www.clontech.com) without disclosure of the donors' personal information. A mouse embryo cDNA library (Cat#: 634209) was purchased from Clontech while a human brain cDNA library (cat# D883001) was purchased from Invitrogen (www.invitrogen.com).

Transient transfection of cells: In many studies, the cDNA and shRNA that were constructed into a pcDNA3.1 expression vector or a retroviral vector were transiently transfected to the desired cells using LipofectAMINE reagent (www.Invitrogen.com) according to the manufacturer's protocol. siRNA oligos were also transfected in the same way.

RT-PCR assay: Total RNA samples were extracted from cultured cells using TRIzol (Invitrogen, Cat. 15596-026) by following the manufacturer's manual. Mouse tissues were homogenized with polytron in a TE buffer containing protease K[69] before RNA extraction with TRIzol. The RNA samples were treated with DNase I to remove DNA residuals, followed by inactivation of the DNase by heating up to 99 ^{0}C for 5 min. A sample of RNA was reverse-transcribed (RT) to cDNA with the TaqMan® reverse transcription kit (Roche Applied Biosystems; www.roche.com) or with M-MLV Reverse Transcriptase (Promega, Cat #. M1705; www.promega.com), following the manufacturers' instruction. The reactions were primed with random hexamers. Commercial RNA samples of normal human tissues were directly reverse-transcribed to cDNA in the same way. cDNA was amplified with PCR which was optimized for each gene and was stopped within the linear portion of the amplification. Each primer pair used in PCR was selected in such a way that the two primers are localized at two different exons with one or several introns in between. In this way if there still is a traceable amount of DNA residual, it either cannot be amplified due to large intron(s) or is amplified as a molecule larger than the expected size. Most primers used are listed in the supplementary table 1. PCR products were separated in agarose gel and visualized with ethidium bromide staining.

DNA sequencing and sequence analyses: The desired band of PCR products in agarose gel was purified and sent to Genewiz, Inc. (www.genewiz.com) for sequencing. Plasmid DNA of desired clones from TOPO or T-A cloning was also sequenced. DNA sequences were analyzed using BLAST of NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) or BLAT of UCSC (http://genome.ucsc.edu/cgi-bin/hgBlat) for alignment, while DNAStar software was used to identify start and stop codons and associated ORF.

3-(4, 5-Dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide (MTT) assay: Cells were seeded in a 96-well micro-plate at 3000 cells per well, five wells per dose or per time point, and incubated at 37^{o}C with 5% CO_2. If the assay was for studying a chemotherapeutic agent, the agent was added 24 hours later at indicated concentrations, with the solvent as the non-treated

control. The culture was continued for the indicated time period, and then MTT was added into each well at a final concentration of 0.5mg/ml followed by incubation at 37^{o}C for three hours in dark. The culture medium containing MTT was discarded and the dye crystals were dissolved in DMSO. The viable cells were detected by reading the absorbance of the metabolized MTT at wavelength of 570nm using the Beckman Coulter AD340 Absorbance Detector (Beckman Coulter Inc, Fullerton, CA).

Cell cycle analysis: Cells were cultured in dishes until they reached 70~80% confluence. For serum starvation study, the medium was replaced with a serum-free one and the cells were allowed on culture for the indicated time period. At cell harvest, the adherent cells were collected, washed with cold PBS, and then fixed overnight with 70% ethanol in PBS at -20^{o}C. The cells were then washed with PBS again and incubated in the dark with in a PBS containing 20 μg/ml propidium iodide (PI) and 200 μg/ml RNase for 30 minutes at room temperature. The cells at different stages of the cell cycle were detected with the BD FACS Calibur Flow Cytometer. Intact cells were gated in the FSC/SSC plot to exclude small debris. The population of cells at different stages of the cell cycle was quantified using ModFit LT software (Verity Software House, Inc., Topsham, ME).

Cell death analyses: Annexin V and PI dual staining was used to visualize and quantify apoptotic and necrotic cells in a procedure similar to cell cycle analysis, but both floating and adherent cells were collected and combined without pre-fixation with 70% ethanol. Apoptotic cells were stained using the Annexin V-FITC Apoptosis Detection Kit (MBL International Corporation, Watertown, MA) following the manufacturer's instruction. All dead cells, i.e. both apoptotic and necrotic cells stained with Annexin V and PI, were analyzed using BD FACS-Calibur Flow Cytometer that placed the apoptosis-detecting Annexin V-FITC signal in FL1 and the necrosis-detecting PI-FITC signal in FL2. Therefore, cells in the upper-left quadrant of the FL1/FL2 dot plot (labeled with PI-FITC only) were necrotic, whereas cells in the bottom-right quadrant were early apoptotic.[28] Cells in the upper-right quadrant (labeled with Annexin V-FITC and PI) were late apoptotic or necrotic.[28] Viable and dead cells were also determined by Acridine orange and ethidium bromide staining as described before,[28] in which cells were seeded in 96-well plates at 3000 cells per well and incubated at 37^{o}C with 5% CO_2. For serum starvation study, the medium was changed to a fresh one without serum when the cells reached roughly 70% confluence, and the cells were allowed on culture for additional days as indicated. Prior to cell staining, the medium was discarded and the cells were washed with PBS first and then incubated with 10 μl PBS containing 10 μg/ml ethidium bromide and 3 μg/ml acridine orange. The cells stained were immediately visualized under the Leica DM IRB inverted fluorescence microscope (Leica Microsystems Inc. Bannockburn, IL). Multiple photos were taken at randomly-selected areas of the well. Photographs were processed using "Image ProPlus 6.0" ultimate image analysis software **(**Media Cybernetics, Inc. Bethesda, MD.

Hematoxylin-eosin and Ki-67 staining and quantification: Xenograft tumors collected from SCID mice were fixed with 10% formalin and embedded in paraffin. The paraffin blocks were sectioned and stained for hematoxylin and eosin. An area of typical lesion was punched out from each block and used to build a tissue array. The tissue array was sectioned and stained for hematoxylin-eosin again. Immunohistochemical staining for Ki-67 was carried out with a rabbit polyclonal Ki-67 antibody from Lab Vision (Cat# RM-9106; www.labvision.com) and with an avidin-biotin complex (ABC) method as described previously.[70] Ki-67 positive and negative tumor cells in five randomly selected areas of each section were counted under light microscope with a 40x lens. The percentage of positive cells in the total number of counted cells was

calculated for further statistical analysis. Stromal areas in the viable parts on each section were estimated subjectively under a grid of a light microscope whereas the stromal areas in necrotic parts were excluded. Necrotic areas were also estimated subjectively in the same way.