SUPPORTING ONLINE MATERIAL

Materials and Methods

Analysis of array CGH/SNP datasets for acute lymphoblastic leukemia and prostate cancer

For Affymetrix SNP arrays, model-based expression was performed to summarize signal intensities for each probe set, using the perfect-match/mismatch (PM/MM) model. For copy number inference, raw copy numbers were calculated for each tumor sample by comparing the summarized signal intensity of each SNP probe set against a diploid reference set of samples. In Agilent two channel array CGH dataset, the differential ratio between the processed testing channel signal and processed reference channel signal was calculated. All resulting relative DNA copy number data were log2 transformed, which reflects the DNA copy number difference between the testing and reference samples. To improve the accuracy of copy number estimation, we employed a reference set normalization method. For each sample, non-sex chromosomes were split into 30Mb region units. The absolute mean of the relative DNA copy number data for the probes from each region was calculated and compared with the other regions. The probes from two regions with minimal absolute mean in each sample were picked up as an internal reference set, representing the chromosomal regions with minimal DNA copy number aberrations. For each sample, log ratios were transformed into a normal distribution with a mean of 0, under the null model assumption for the reference probe set. The normalization method was implemented by perl programming.

Amplification Breakpoint Ranking and Assembly (ABRA)

In order to nominate partially amplified gene fusions systematically from genomic data, we employed ABRA across a compendium of data from cancer cell lines, as breakpoint analyses are more reliable in uniform cellular populations as opposed to tumors which are made up of multiple cell types many of which are not malignant. The workflow is described in **Supplementary Fig. 1c**.

First, the copy number data from the array CGH or array SNP datasets were segmented by the circular binary segmentation (CBS) algorithm [1]. The level of amplification was determined by comparing the relative copy number data of the amplifications with the neighboring segments, and the breakpoints having equal to or more than 2 copies number gain were selected ($\geqslant$0.75). Amplifications spanning more than 500kb are included in the analysis. The genomic position of each amplification breakpoint was mapped with the genomic regions of all human genes. The genomic region of each human gene was designated as the starting of the transcript variant most approaching the 5' of the gene, and the end of the variant most approaching the 3' of the gene. The partially amplified genes were classified into candidate 5' and 3' partners based on the association of amplification breakpoints with gene placements. 5' amplified genes are considered as 5' partners, 3' amplified genes as 3' partners.

Second, we sought to identify the partially amplified "cancer genes" as driver fusion gene candidates. This can be easily achieved by mapping 3' amplified genes to known cancer genes defined by cancer gene census (http://www.sanger.ac.uk/genetics/CGP/Census/)[2], which however, may overlook the less characterized ones. To evaluate the relevance of partially amplified genes underlying cancer, we adopted the "concept signature technology" (ConSig) developed in our previous study [3], which can preferentially identify biologically meaningful genes based on their association with the "molecular concepts" frequently found in known cancer genes. This score is especially discriminative for 3' fusion genes [3]. We therefore rated the 3' amplified genes with acceptable breakpoints (see below criteria, **Supplementary Fig. 2a**), by their radial concept signature scores (in brief ConSig Score). The

top scored 3' amplified cancer genes were considered as driver fusion gene candidates.

Third, the level of amplification for the selected 3' amplified gene was matched with 5' amplified genes from the same cell line to nominate putative 5' partners. The actual location and the quality of the breakpoint will be manually curated with the un-segmented relative quantification of DNA copy number data. The situations when the amplification breakpoint is not acceptable (**Supplementary Fig. 2**):

(1) Multiple intragenic breakpoints;
(2) The candidate is not the gene closest to the amplification breakpoint;
(3) The amplification starts from existing copy number increase and the breakpoint is not sharp;
(4) The breakpoint locates at the centromere or the end of the chromosome;
(5) The breakpoint is the result of a small deletion within an amplification;
(6) The breakpoint is found in a majority of samples.

A major concern for the breakpoint analysis was that the segmentation process could have slightly different estimation of the breakpoints from the actual location. This is especially critical for breakpoint assembling. To overcome this problem, the DNA breakpoints within 10 kb up and 1kb downstream region of a gene will be assigned to this gene during breakpoint ranking; and 20kb up- and downstream during breakpoint assembling. In practice, this window can be adjusted to improve the performance of ABRA analysis.

In this study, we applied *in silico* amplification breakpoint assembly to nominate the fusion partners. In general, we recommend applying this approach to cancer cell lines that have uniform cell populations, thus the copy number estimation will be more reliable. And repeated hybridization using the highest resolution microarray CGH or SNP platforms are usually needed to pinpoint the intragenic amplification breakpoints. Alternatively, the amplification breakpoints can be assembled by analyzing the samples harboring the partially amplified fusion genes with paired-end transcriptome sequencing. This strategy is particularly useful to nominate fusion partners in tissue samples. Further, the link of ABRA with next generation sequencing will comprise a highly cost-effective approach to nominate causal genetic aberrations from this data. By leveraging the public or private cancer genomic datasets, we can nominate candidate amplified fusion genes, and then focus the sequencing effort to a small number samples harboring these candidates. This approach will be particularly valuable considering the exponential accumulation of public genomic datasets from large cancer genome projects, such as The Cancer Genome Atlas (TCGA), the Tumor Sequencing Project (TSP), the Cancer Genome Project (CGP), and individual deposits by laboratories world-wide.

Cell lines

The benign immortalized prostate cell line RWPE, prostate cancer cell line DU145, PC3, Ca-HPV-10, WPE1-NB26 and NCI-H660, Fibroblast cell line NIH 3T3, and human embryonic kidney cell line HEK were freshly obtained from the American Type Culture Collection (Manassas, VA). VCaP was derived from a vertebral metastasis from a patient with hormone-refractory metastatic prostate cancer 4, and was provided by K. Pienta. The prostate cancer cell lines C4-2B, LAPC4 and MDA-PCa 2B were provided by E. Keller. The prostate cancer cell line 22-RV1 was provided by J. Macoska. Key cell lines used for all molecular biology and functional studies were obtained from ATCC and were authenticated by manufacturer through analysis of informative polymorphic markers (short tandem repeat analysis), cell morphology and species authentication. For the ancillary cell lines provided by collaborators, no additional authentication was done by authors. These cell lines were not used for functional studies in the context of the Kras fusion only as controls.

Microarray comparative genomic hybridization (array CGH)

To nominate potential driver gene fusions in prostate cancer cell lines, we profiled ten prostate cancer cell lines on Agilent-014698 Human Genome CGH Microarray 105A (Agilent Technologies, Palo Alto, CA), including 22RV1, C4-2B, CA-hpv-10, DU145, LAPC4, MDAPCa-2b, NCI660, PC3, VCaP, and WPE1-NB26. All cell lines were grown in full serum in accordance with the distributor's instructions. The genomic DNA extracted from those cell lines were hybridized against reference human male genomic DNA (6 normal individuals, Promega, #G1471) to oligonucleotide printed in the array format according to manufacture's protocol. Analysis of fluorescent intensity for each probe will detect the copy number changes in cancer cell lines relative to normal reference genome (Genome build 2004). Replicate array CGH hybridizations of DU145 were done to nominate 5' partners of *KRAS*. The array CGH data used in this study have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus with the accession number GSE26447.

Paired-end transcriptome sequencing and analysis

DU145 mRNA samples were prepared for sequencing using the mRNA-seq sample prep kit (Illumina) following manufacturers protocols. The raw sequencing image data were analyzed by the Illumina analysis pipeline, aligned to the unmasked human reference genome (NCBI v36, hg18) using the ELAND software (Illumina). The paired reads were then analyzed as previously described to nominate mate-pair chimeras [5].

Reverse-transcription PCR (RT-PCR) and sequencing

Total RNA from all samples was isolated with Trizol (Invitrogen, Carlsbad, CA) according to the manufacturer's protocols. RNA integrity was verified by Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA). Complimentary DNA was synthesized from one microgram of total RNA, using SuperScript III (Invitrogen, Carlsbad, CA) or high-capacity cDNA reverse transcription kit (Applied Biosystems, Foster City, CA) in the presence of random primers. The reaction was carried out for 60 minutes at 50ºC and the cDNA was purified using microcon YM-30 (Millipore Corp, Bedford, MA, USA) according to manufacturer's instruction and used as template in PCRs. All oligonucleotide primers used in this study were synthesized by Integrated DNA Technologies (Coralville, IA) and are listed in **Supplementary Table 1**. Polymerase chain reaction was performed with Platinum Taq High Fidelity and fusion-specific primers for 35 cycles. Products were resolved by electrophoresis on 1.5% agarose gels, and bands were excised, purified and TOPO TA cloned into pCR 4-TOPO TA vector (Invitrogen, Carlsbad, CA). Purified plasmid DNA from at least 4 colonies was sequenced bi-directionally using M13 Reverse and M13 Forward primers on an ABI Model 3730 automated sequencer at the University of Michigan DNA Sequencing Core.

Quantitative PCR (qPCR)

Quantitative PCR (qPCR) was performed using the StepOne Real Time PCR system (Applied Biosystems). Briefly, reactions were performed with SYBR Green Master Mix (Applied Biosystems) cDNA template and 25 ng of both the forward and reverse fusion primers q1/q2 using the manufacturer recommended thermocycling conditions. For each experiment, threshold levels were set during the exponential phase of the QPCR reaction using the StepOne software. The amount of each target gene relative to the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) for each sample was determined using the comparative threshold cycle (Ct) method (Applied Biosystems User Bulletin #2, http://docs.appliedbiosystems.com/pebiodocs/04303859.pdf). For the experiments presented in **Figure 1c**, the relative amount of the target gene was calibrated to the relative amount

from a benign prostate.

Nuclease protection assay

Ribonuclease protection assays were performed utilizing a 230 bp fragment spanning the UBE2L3-KRAS fusion junction. An *in vitro* antisense transcript was labeled to a specific activity of 3 X 10^8 cpm per microgram using ^{32}P UTP and T7 RNA polymerase according to the conditions specified in the T7 MaxiScript protocol (Applied Biosystems). 20 micrograms of sample RNAs were hybridized to the labeled probe and digested according to the conditions specified in the RPA III protocol (Applied Biosystems). Digested samples were electrophoresed on 7% denaturing polyacrylamide gels, using a sequencing ladder as a size marker. Following electrophoresis, the gels were dried and exposed to storage phosphor screens and imaged with a Molecular Imager (BioRad).

Fluorescence in situ hybridization (FISH)

To evaluate the fusion of *UBE2L3* with *KRAS*, we employed a two-color, two-signal FISH strategy, with probes spanning the respective gene loci. The digoxin-dUTP labeled BAC clone RP11-317J15 was used for the *UBE2L3* locus and the biotin-14-dCTP BAC clone RP11-608F13 was used for the *KRAS* locus. To detect possible translocations at *KRAS* locus, we used a break-apart FISH strategy, with two probes spanning the *KRAS* locus (digoxin-dUTP labeled BAC clone RP11-68I23, (5' *KRAS*) and biotin-14-dCTP labeled BAC clone RP11-157L6 (3' *KRAS*)). All BAC clones were obtained from the Children's Hospital of Oakland Research Institute (CHORI). Prior to FISH analysis, the integrity and purity of all probes were verified by hybridization to metaphase spreads of normal peripheral lymphocytes.

For interphase FISH on DU145 cells, interphase spreads were prepared using standard cytogenetic techniques. For interphase FISH on a series of prostate cancer tissue microarrays, tissue hybridization, washing and color detection were performed as described [6,7]. The total evaluable cases for *KRAS* split probes on the tissue microarrays include 78 localized and 25 metastatic prostate tumors (University of Michigan cohort). For evaluation of the interphase FISH on the TMA, an average of 50-100 cells per case were evaluated for assessment of the *KRAS* rearrangement. In addition, formalin- fixed paraffin-embedded (FFPE) tissue sections from the two index cases (PCA0211, PCA0216) were used to validate the findings.

Western Blotting

The prostate cancer cell lines DU145 were transfected with siRNA duplex (Dharmacon, Lafayette, CO, USA) against UBE2L3 (5'-CCACCGAAGATCACATTTA-3'), KRAS (5'-GAAGTTATGGAATTCCTTT-3') or the fusion junction (5'-CCGACCAAGGCCTGCTGAA-3') by oligofectamine (Invitrogen). DU145 transfected with non-targeting siRNA and RWPE cells were used as negative control. Post 48 hours transfection, cells were homogenized in NP40 lysis buffer (50 mM Tris-HCl, 1% NP40, pH 7.4, Sigma, St. Louis, MO), and complete proteinase inhibitor mixture (Roche, Indianapolis, IN). Ten micrograms of each protein extract were boiled in sample buffer, separated by SDS-PAGE, and transferred onto Polyvinylidene Difluoride membrane (GE Healthcare, Piscataway, NJ). The membrane was incubated for one hour in blocking buffer [Tris-buffered saline, 0.1% Tween (TBS-T), 5% nonfat dry milk] and incubated overnight at 4ºC with the following antibodies: anti-RAS mouse monoclonal (1:1000 in blocking buffer, Millipore Cat #: 05-516), anti-KRAS rabbit polyclonal (1:1000, Proteintech Group Inc., Cat #: 12063-1-AP) and anti-beta Actin mouse monoclonal (1:5000, Sigma Cat #: A5441) antibodies. Following three washes with TBS-T, the blot was incubated with horseradish peroxidase-conjugated secondary antibody and the signals

visualized by enhanced chemiluminescence system as described by the manufacturer (GE Healthcare). To test fusion protein expression in multiple prostate derived cell lines, lysates from DU145, PrEC, RWPE, 22RV1, VCaP, PC3 either untreated or treated with 500nM bortezomib for 12 hours were used. Bortezomib treated HEK cells over-expressing UBE2L3-KRAS fusion protein was used as positive control.

To explore the activation of MAPK signaling pathways, protein lysates from NIH 3T3 stable cell lines expressing UBE2L3-KRAS, V600E mutant BRAF, G12V mutant KRAS, and vector controls were probed with phospho MEK1/2, phospho p38 MAPK, phospho Akt, and equal loading was demonstrated by probing for the respective total proteins and beta Actin. For ERK activation analysis NIH 3T3 cells were starved for 12 hours before the immunoblot analyses of phospho and total erk 1/2. All antibodies for the MAPK signaling proteins were purchased from Cell Signaling Technologies.

Multiple Reactions Monitoring Mass Spectrometry

Du145, LnCaP and VCaP cells were grown to 70% confluence and treated with bortezomib. After 24 hours, cells were harvested and whole cell protein lysates were prepared in RIPA buffer (Pierce Biotechnology, Rockford, IL, USA) with the addition of protease inhibitor complete mini cocktail (Roche, Indianapolis, IN, USA). Lysates were cleared by centrifugation and separated by SDS PAGE (Novex, 18 % Tris-Glycine, Invitrogen, Carlsbad, CA, USA). 12 equal sized bands from 15-40 kDa regions were excised for in-gel trypsin digestion. Lyophilized peptides from each gel slice were re-suspended in 3 % acetonitrille, 0.1 % formic acid containing 25 fmol of each stably isotopically labeled peptide internal standards (Sigma-Aldrich Corp.St. Louis, MO, USA). Peptides were then separated and measured by CHIP HPLC-multiple reaction monitoring mass spectrometry (MRM-MS). Three transitions for each stably isotopically labeled internal standard and three transitions for endogenous peptides were measured. An overlap of all 6 transitions for each peptide in retention time indicated a positive measurement.

Immunoprecipitation

For the detection of endogenous ubiquitinated UBE2L3-KRAS, fifty millions of DU145 cells were lysed in RIPA buffer containing 5mM N-ethylmaleimide (NEM) and subjected to immuno-precipitation (IP) with rat monoclonal anti-RAS antibody conjugated agarose (Santa Cruz, SC34AC). The beads were extensively washed with RIPA buffer and 1M NaCl containing PBS. The bound proteins were eluted with PBS containing 2% SDS and resolved on 4-12% SDS-PAGE gradient gel and analyzed with the indicated antibodies. For 293T experiments, expression plasmids of HA-ubiquitin and UBE2L3-KRAS/control vector were co-transfected using Fugene (Invitrogen) for 48 hours. The cells were lysed and exogenously expressed UBE2L3-KRas was immunoprecipitated with anti-rat Ras monoclonal antibody (Santa Cruz, clone 238) and washed with RIPA buffer. Then immunoblotting was done using mouse anti-pan RAS (EMD CHEMICALS, Ab-3) and anti-HA (Covance, clone 16B12) antibodies.

In vitro overexpression of the UBE2L3-KRAS fusion

Expression plasmids for UBE2L3-KRAS were generated with the pDEST40 (with or without 5' FLAG) and pLenti-6 vectors (without 5'FLAG). NIH 3T3 cells were maintained in DMEM with 10%FBS and transfected with either the pDEST40 plasmid or pDEST40 construct containing the UBE2L3-KRAS open reading frame using Fugene 6 transfection reagent (Invitrogen). After three days, transfected cells were selected using 500ug/ml Geneticin. After three weeks of selection stable cell lines were established for both the vector and UBE2L3-KRAS fusion, and were used for further

analyses. Constructs for the G12V mutant KRAS (Addgene plasmid 9052), V600E mutant BRAF (Addgene plasmid 15269), and their respective pBABE–puro vector (Addgene plasmid 1764) were obtained from Addgene (Cambridge, MA, USA). These plasmid constructs were transfected in NIH 3T3 cells maintained in 10% calf serum and stable lines were generated using puromycin 1ug/ml for selection. These stable cell lines were used as controls for immunoblot analysis of the RAS-MAPK signaling pathways.

To overexpress UBE2L3-KRAS fusion in the prostate derived normal cell lines, lentiviral particles expressing the UBE2L3-KRAS fusion were transfected in RWPE cells (pLenti-6 vector). Three days after infection, individual clones were selected using 3ug/ml blasticidin. Four independent clones were analyzed for their increased growth potential. Both the NIH 3T3 and the RWPE overexpression models were tested for UBE2L3-KRAS fusion by qPCR (Supplementary Fig. 7b-c) and Western blotting.

Cell proliferation assay

For cell proliferation analysis, 10,000 cells of NIH 3T3 expressing UBE2L3-KRAS fusion or the vector were plated on 24 well plates in duplicate wells and cell counts were performed using a Coulter Counter (Beckman Coulter, Fullerton, CA) at the indicated times. Similar assays were performed using RWPE stable clones expressing UBE2L3-KRAS fusion or vector. Both cell proliferation assays were performed twice and data from representative assays are presented. Effect of fusion and control shRNAs on Du145 cell proliferation was monitored using WST-1 assay (Roche).

Basement Membrane Matrix Invasion assay and MEK inhibition

Stable pooled population of RWPE cells over-expressing vector pLenti-6, fusion transcript UBE2L3-KRAS, wild type WT-KRAS or oncogenic KRAS-G12V were seeded onto a matrigel coated transwell inserts (BD Biosciences) in the presence of DMSO or MEK inhibitor (U0126; 20μM) and processed as the manufacturer's recommendation. After 48 hours the inserts were fixed in 70% ethanol and stained with crystal violet. Each transwell insert was detained in 100μl of 10% acetic acid, and quantified by colorimetric assay at absorbance 560nm. DU145 was used as a control for the invasion assay. The invasion assay was also performed for RWPE stable clones over-expressing UBE2L3-KRAS or vector pLenti-6 and DU145 without any treatment.

Foci Formation Assay

Transfections were performed using Fugene 6 according to the manufacturer's protocol (Roche Applied Sciences). NIH 3T3 cells (1.5×10^5) in 35-mm plastic dishes were transfected with 2 μg of DNA of the plasmid of interest. Plasmids for fusion transcript UBE2L3-KRAS and oncogenic KRAS G12V were used along with control plasmids (pDEST40 and pBABE respectively). Three days after transfection, cells were split into one 140-mm dishes containing DMEM with 5% calf serum (Colorado Serum Company). The cultures were fed every 3-4 days. After 3 weeks, the cells were stained with 0.2% crystal violet in 70% ethanol for the visualization of foci, and were counted on colony counter (Oxford Optronix Ltd., Oxford UK, software v4.1, 2003). Counts were further confirmed manually.

FACS Cell Cycle Analysis

Propidium iodide–stained stable NIH 3T3 cells expressing the UBE2L3-KRAS fusion or vector were analyzed on a LSR II flow cytometer (BD Biosciences, San Jose, CA) running FACSDivia, and cell cycle phases were calculated using ModFit LT (Verity Software House, Topsham, ME).

Stable Knockdown of UBE2L3-KRAS fusion in DU145 cells

For stable knockdown of fusion KRAS, shRNA targeting UBE2L3-KRAS was obtained from System Biosciences (Mountain View, CA). Lentiviruses were generated by the University of Michigan Vector Core. Prostate cancer cell line DU145 were infected with shRNA UBE2L3-KRAS lentiviruses or vector only, and stable cell lines were generated by selection with puromycin (Invitrogen, Carlsbad, CA).

NIH 3T3 and RWPE -UBE2L3-KRAS and DU145-shUBE2L3-KRAS Xenograft Model

Four weeks old male Balb C nu/nu mice were purchased from Charles River, Inc. (Charles River Laboratory, Wilmington, MA). Stable NIH 3T3 and RWPE cells over expressing fusion transcript UBE2L3-KRAS or NIH 3T3-Vector (2×10^6 cells), and pooled or single clone population of Du145 cells with the stable knockdown of fusion transcript UBE2L3-KRAS (5×10^6 cells) were resuspended in 100µl of saline with 20% Matrigel (BD Biosciences, Becton Drive, NJ) and were implanted subcutaneously into the left or both left and right flank regions of the mice. Mice were anesthetized using a cocktail of xylazine (80-120 mg/kg IP) and ketamine (10mg/kg IP) for chemical restraint before implantation. Eight mice were included in each group. Growth in tumor volume was recorded everyday by using digital calipers and tumor volumes were calculated using the formula $(\pi/6)$ $(L \times W2)$, where L = length of tumor and W = width. All procedures involving mice were approved by the University Committee on Use and Care of Animals (UCUCA) at the University of Michigan and conform to their relevant regulatory standards.

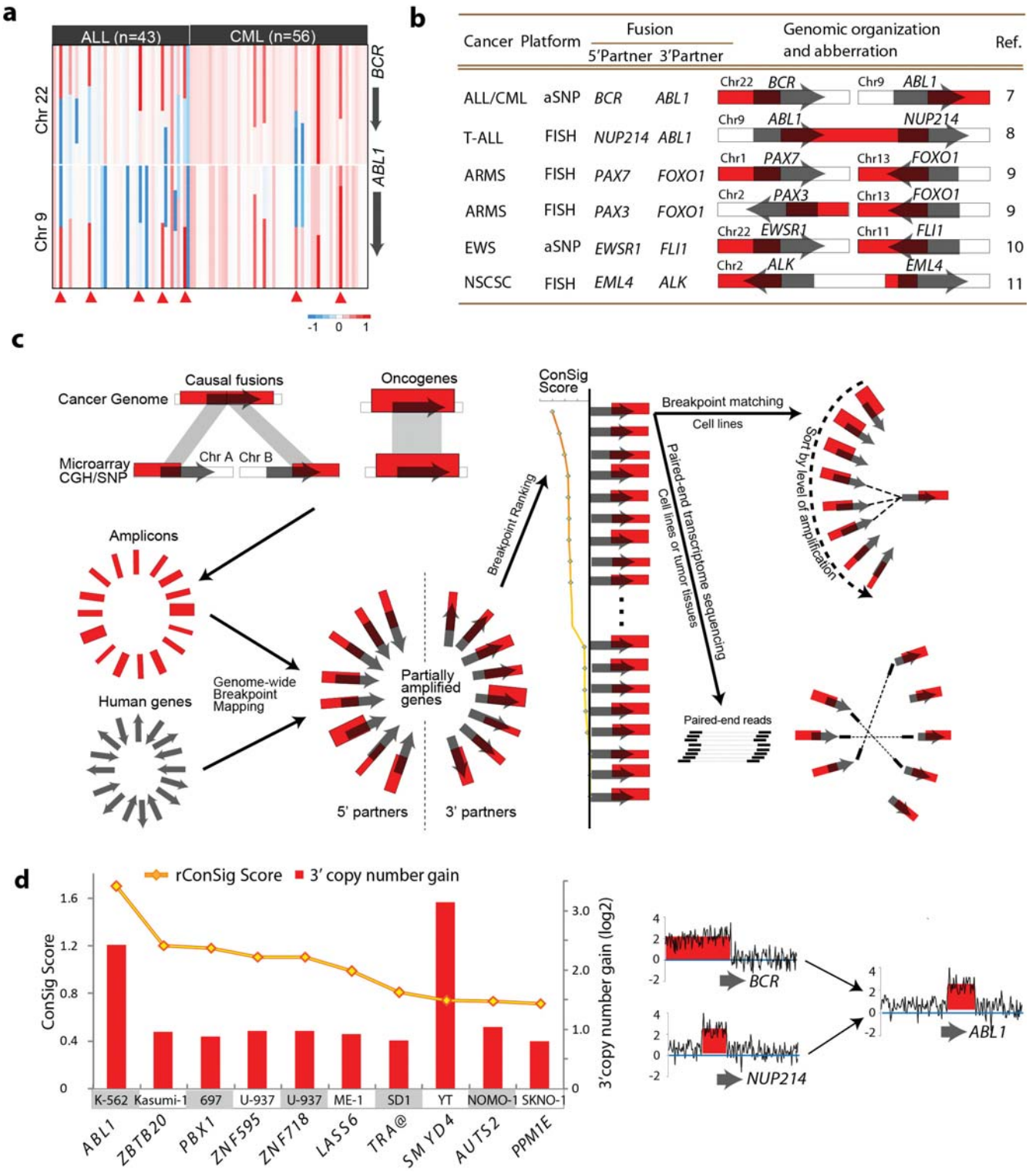
Chicken embryo tumor formation assay

Chicken embryo tumor formation assay was performed by inoculating 2 million UBE2L3-KRAS stable knockdown DU145 cells on the CAM of 10-day old chicken embryo. On day 18, tumors were exercised and weighed. Scramble shRNA stable DU145 cells was used as control.

Immunofluorescence and confocal microscopy

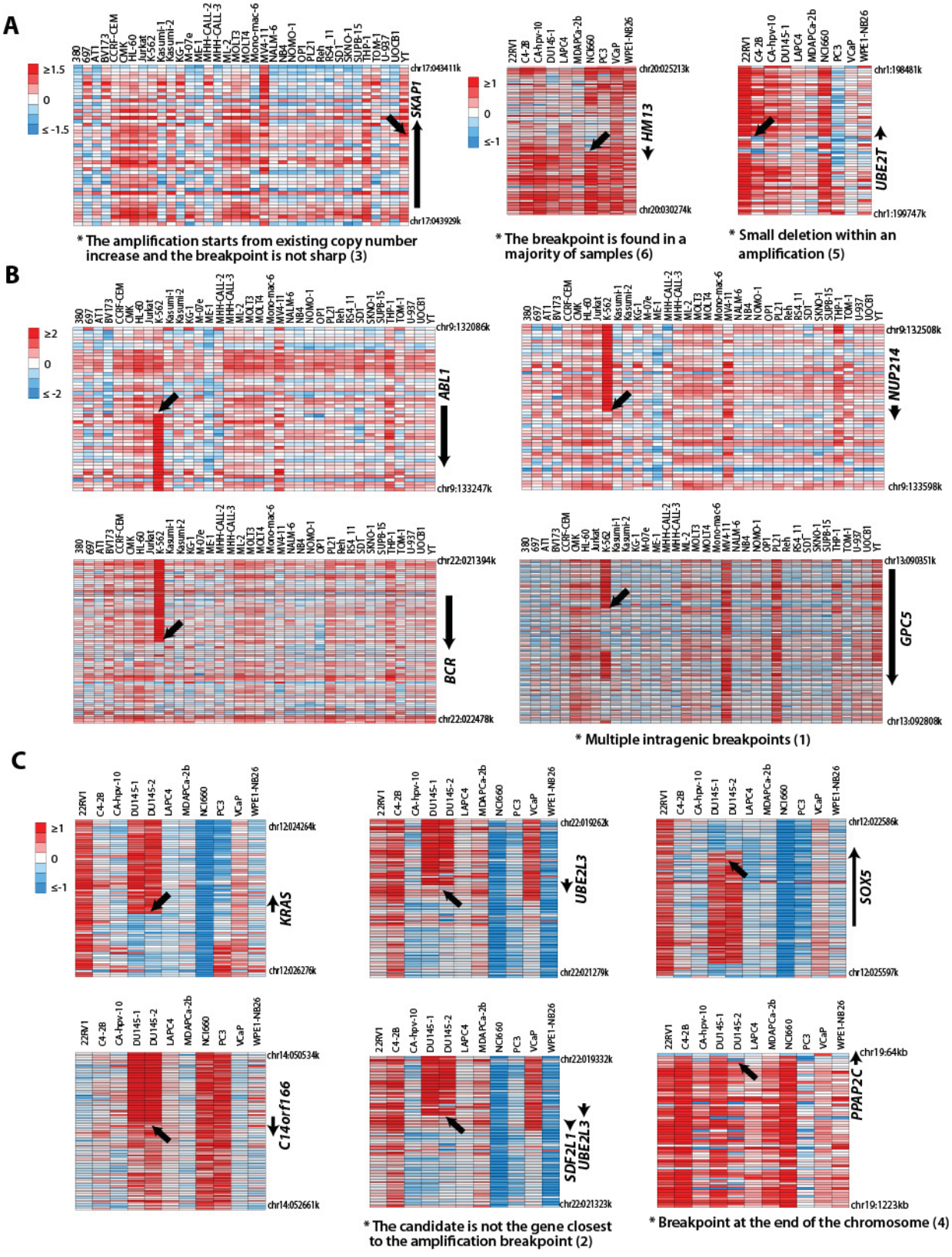
The stable cell lines are fixed in cold methanol and blocked in PBS-Tween 20 containing 5% normal donkey serum for 1 hour. A mixture of mouse anti-KRAS monoclonal antibody (Invitrogen, Cat #: 415700) and rabbit anti-LAMP1 polyclonal antibody (Santa Cruz Biotechnology, Cat #: sc-5570) was added to the slides at 1:50 and 1:20 dilutions, respectively, and incubated overnight at 4℃ following standard immunofluorescence methods[8]. The results were reconfirmed using two additional antibodies: mouse anti-KRAS monoclonal antibody (1:50 in blocking buffer, Sigma-Aldrich, Cat#: WH0003845M1) and anti-RAS monoclonal (1:20 in blocking buffer, Millipore Cat #: 05-516).

Supplementary Figures



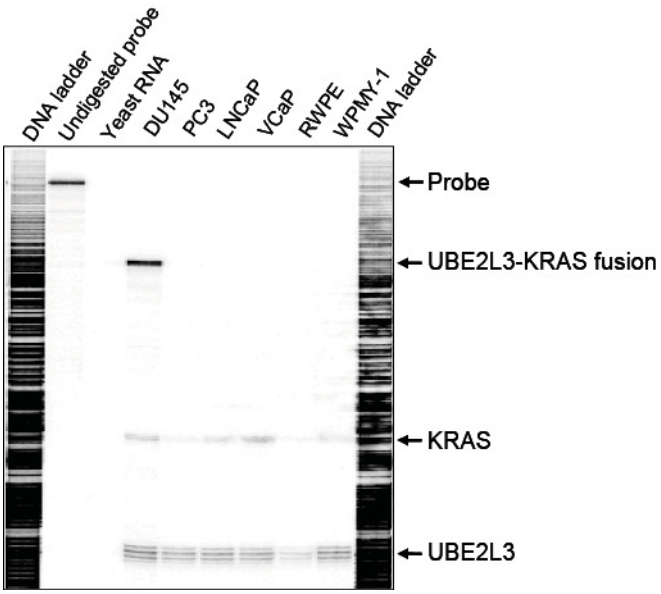
Supplementary Figure 1. The principle and workflow of amplification breakpoint ranking and assembly analysis. (a) The recurrent *BCR-ABL1* fusion was often amplified in fusion positive leukemia patients. Red arrows heads denote the leukemia samples with amplification of the 5' region of *BCR* and 3' region of *ABL* resulting from amplification of *BCR-ABL1* fusion. The acute lymphoblastic

leukemia (ALL) samples presented in the figure are all BCR-ABL1 fusion positive, while chronic myelogenous leukemia (CML) is a tumor entity defined by *Philadelphia chromosome*. Color scales indicate $\log_2$ transformed relative DNA copy number. **(b)** Table displaying known recurrent gene fusions which are accompanied by characteristic focal amplifications in a subset of patients. These amplifications tend to affect 5' end of 5' partners, and 3' end of 3' partners. The fusion partners are indicated by grey arrows with the amplified regions colored red. Left or right arrows indicate - and + DNA strands, respectively. **(c)** The bioinformatics workflow of Amplification Breakpoint Ranking and Assembly analysis. The fusion partners are indicated by grey arrows with the amplified regions colored red. The level of amplification is depicted by the height of the amplified region. This pipeline incorporates a whole genome breakpoint mapping analysis, which enables genome-wide detection and mapping of amplification breakpoints and thus assembly of the fusion partners. The *in silico* assembly of fusion partners can be applied to cancer cell lines with uniform cell populations. Paired-end transcriptome sequencing provides an alternative way to link the fusion partners in both cancer cell lines and tissues. **(d)** Left panel, amplification breakpoint analysis and ConSig scoring of 3' amplified genes from 36 leukemia cell lines identify *ABL1* as a fusion gene associated with 3'amplification. The ConSig scores are depicted by the yellow line and the level of 3' amplification for each 3' fusion gene candidate is depicted by red columns. Right panel, matching the amplification level of 5' amplified genes in K562 cells with *ABL1,* nominates *BCR* and *NUP214* as 5' partner candidates. The relative quantification of DNA copy number data from the genomic regions 1Mb apart from the candidate fusion genes is shown. The x axis indicates the physical position of the genomic aberrations. The fusion partners are indicated by grey arrows.

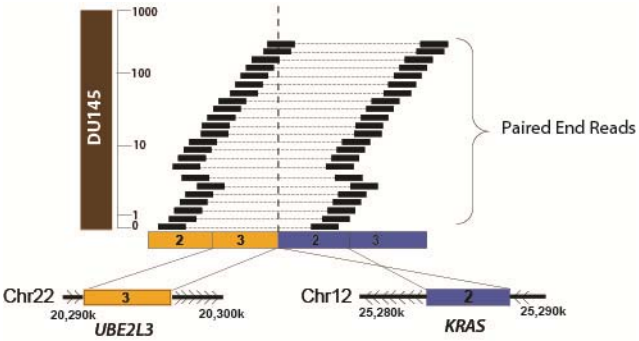


Supplementary Figure 2. SNP array and array CGH data for representative 5' and 3' fusion partner candidates (genes with 5' or 3' amplification) depicting the criteria of manual curation. (a) The relative DNA copy number data for representative candidate 3' partners in leukemia and prostate cancer cell lines with unacceptable breakpoints. (b) The array CGH data for candidate 5' fusion

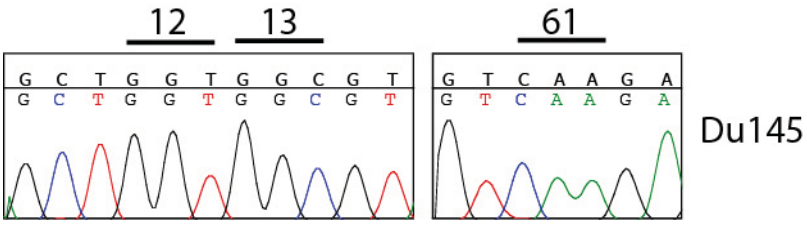
partners of *ABL1* identified by amplification breakpoint assembling analysis on K-562, together with other leukemia cell lines. (c) The array CGH data for candidate 5' fusion partners of KRAS on DU145 (two replicate hybridizations) and other prostate cancer cell lines. The log 2 transformed relative copy number data were depicted by color scales. Amplifications are colored red, and deletions blue. *The situations when the amplification breakpoint is not acceptable are detailed in the methods section.



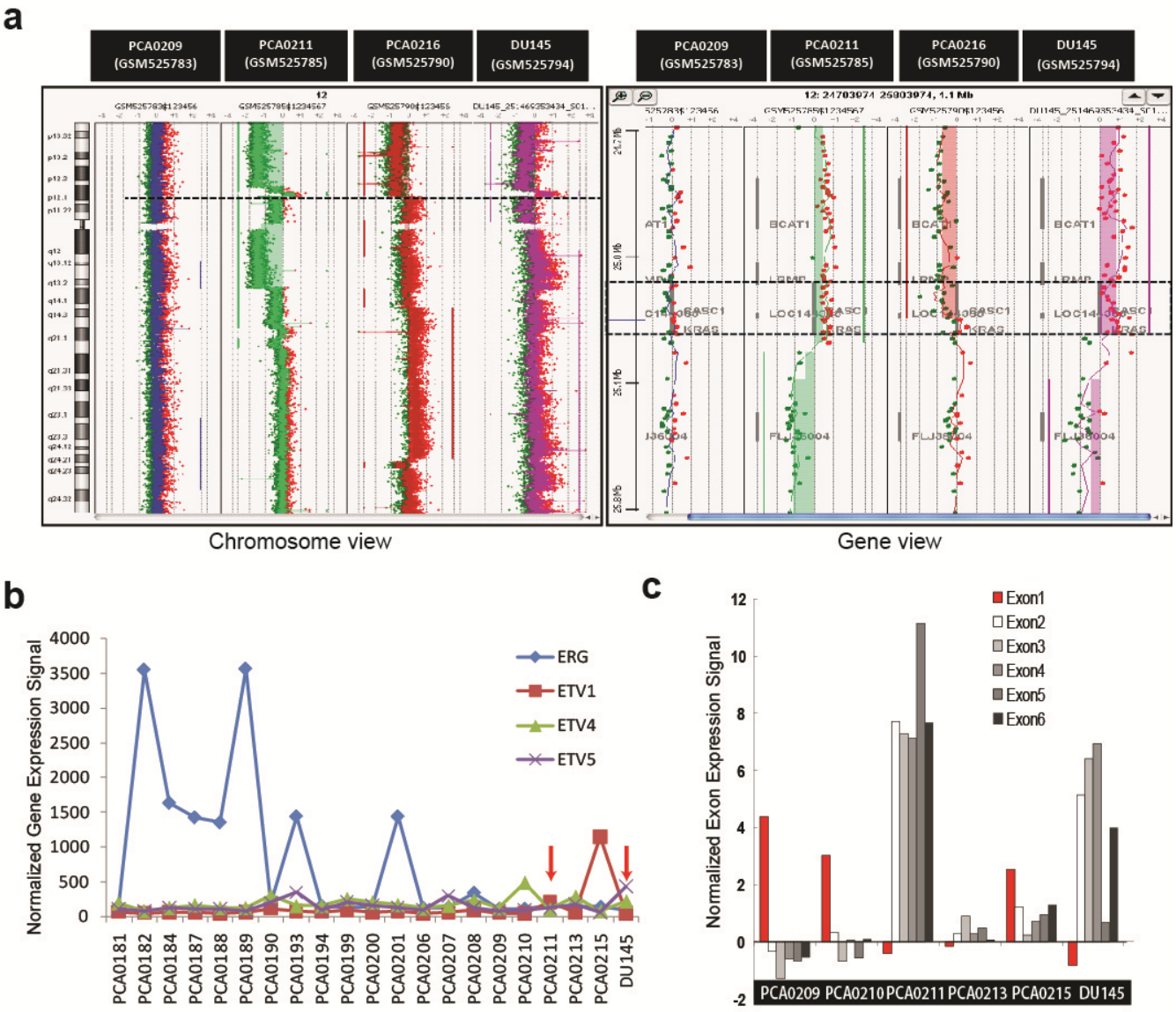
Supplementary Figure 3. RNase Protection Assay revealed abundant UBE2L3-KRAS fusion RNA in DU145 cells but not in other prostate cancer cell lines. Ribonuclease protection assays were performed using a 230 bp fragment spanning the UBE2L3-KRAS fusion junction. A sequencing ladder was used as a size marker. Yeast RNA was used as negative control.



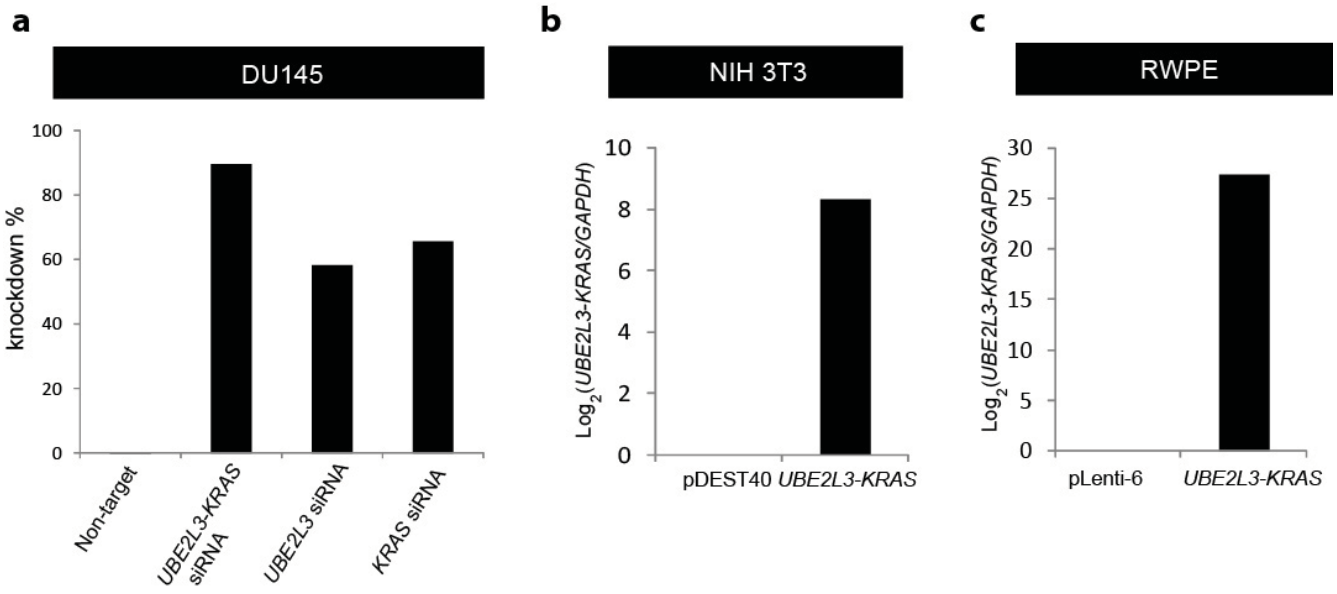
Supplementary Figure 4. Schematic of paired-end sequencing coverage of the fusion between *UBE2L3* (orange) and *KRAS* (blue) in DU145. Frequency of mate pairs is shown in log scale. Chromosomal context of the fusion genes is represented by colored bars punctuated with black lines.



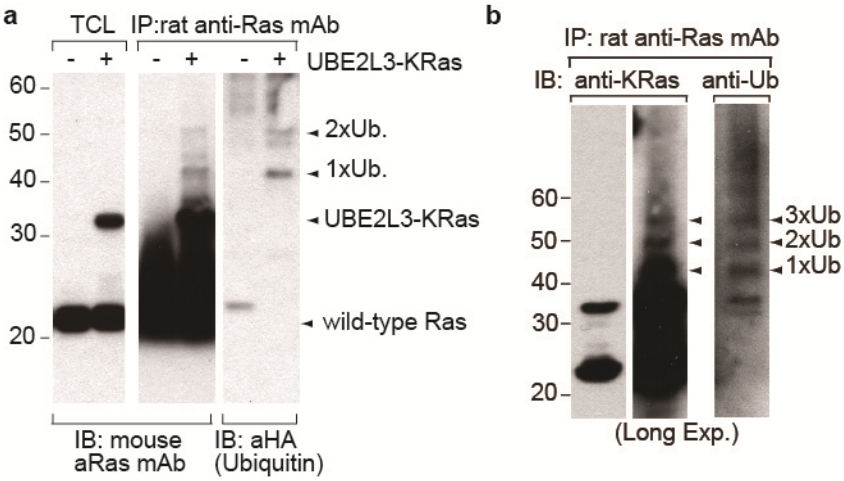
Supplementary Figure 5. Analysis of the fusion sequences from DU145 did not reveal canonical mutation in KRAS fusion allele. Line graphs show the Sanger sequencing results of the fusion transcripts at the positions of KRAS codon 12, 13, 61 from DU145 cells.



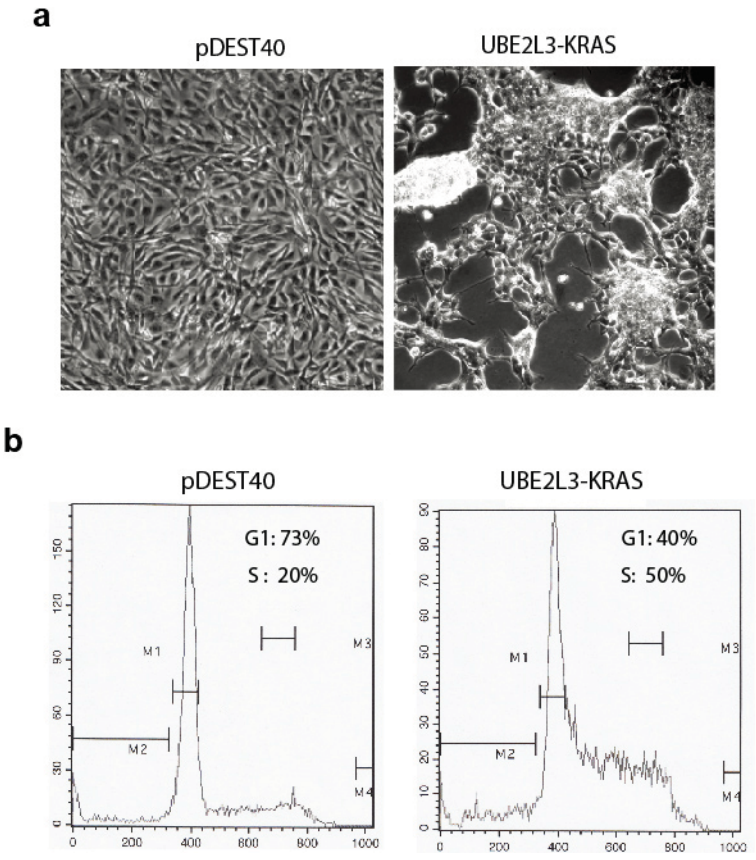
Supplementary Figure 6. Recurrent rearrangements of KRAS in metastatic prostate tumors from Memorial Sloan-Kettering Cancer Center. (a). Comparative genomic hybridization analysis of KRAS locus showing normal copy number in GSM525783 (PCA0209), an intragenic deletion in GSM525790 (PCA0216), and partial amplifications in GSM525785 (PCA0211) and DU145 cell line. **(b)**. Microarray gene expression data for PCA0211 and DU145 showing no expression of ETS family genes (ERG/ETV1, 4, 5). **(c)**. Exon array expression data for PCA0211 and DU145 showing over-expression of KRAS Exon 2-6 in contrast to Exon 1 in both Du145 and PCA0211. The expression signal of each exon is normalized based on its average expression signal and the standard deviation across 185 samples profiled in this dataset (GSE21034).



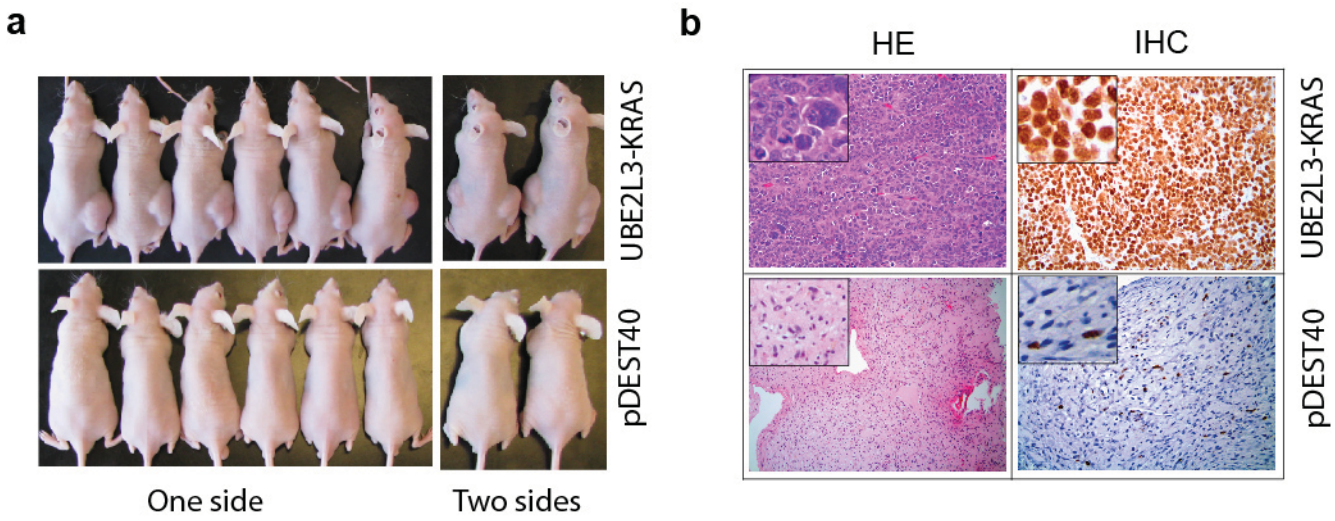
Supplementary Figure 7. qPCR confirmation of siRNA knockdown and ectopic expression of *UBE2L3-KRAS* fusion. (a) qPCR confirmation of *UBE2L3-KRAS* knockdown by siRNA against the fusion junction, wild-type UBE2L3, and wild-type KRAS on DU145. Percent knockdown of the fusion transcript is indicated. Non-targeting siRNA was used as control. (b-c), qPCR confirmation of NIH 3T3 and RWPE cells expressing UBE2L3-KRAS fusion. NIH 3T3 cells **(b)** were transfected with the empty pDEST40 vector or the UBE2L3-KRAS fusion. RWPE cells **(c)** were transfected with lentiviral particles harboring the empty pLenti-6 vector or UBE2L3-KRAS fusion. The relative amount of the fusion transcript was estimated using GAPDH as internal control.



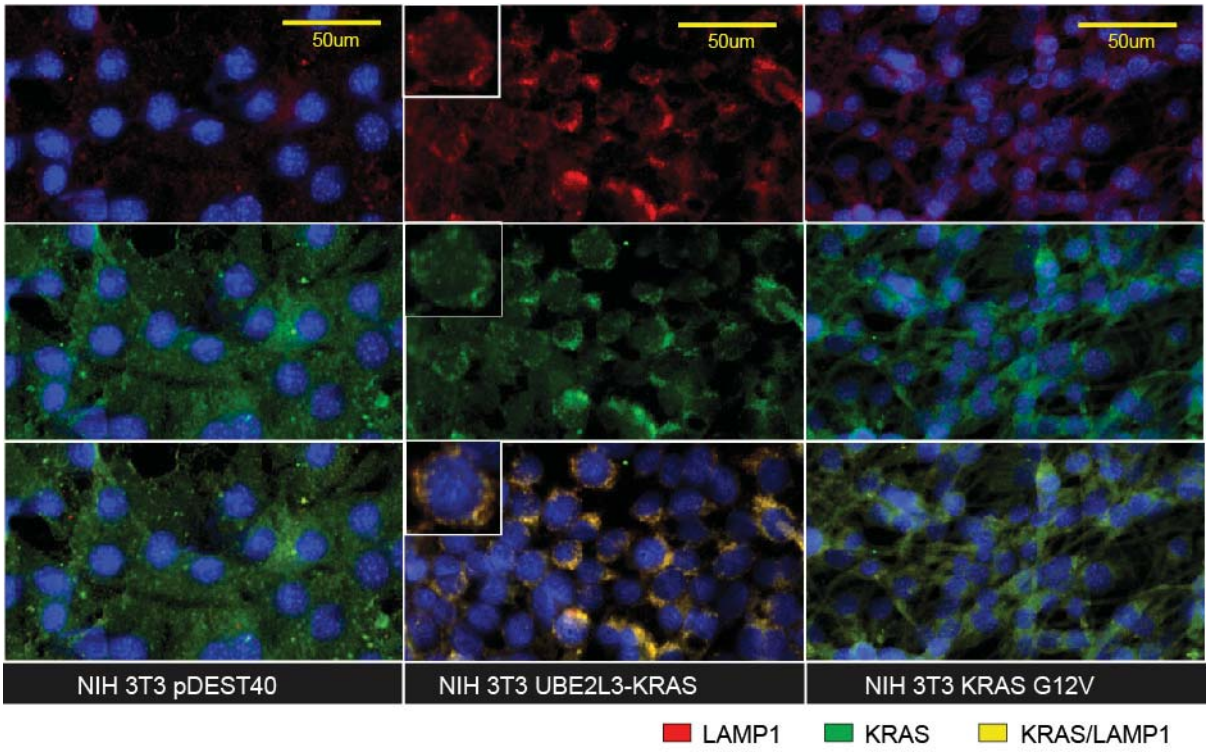
Supplementary Figure 8. Detection of ubiquitinated UBE2L3-KRAS in HEK 293T and DU145 cells. (a) Expression plasmids of HA tagged ubiquitin and UBE2L3-KRAS or control vector were co-transfected into 293T cells. Endogenous Ras and ectopically expressed UBE2L3-KRAS were immunoprecipitated with anti-rat Ras monoclonal antibody. The immunoblotting using anti-pan Ras and anti-HA antibodies revealed ubiquitination ladder of UBE2L3-KRAS around 40~50kDa range, which were not seen in UBE2L3-KRAS non-transfected cells. (**b**) Endogenous Ras and UBE2L3-KRAS were immunoprecipitated with anti-rat Ras monoclonal antibody from DU145 cells. Ubiquitinated UBE2L3-KRAS fusion proteins were detected by western blotting using anti-KRas and anti-ubiquitin antibodies. TCL, total cell lysate.



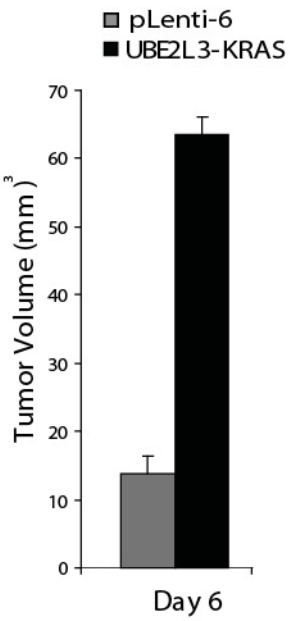
Supplementary Figure 9. The transforming activity of UBE2L3-KRAS in NIH 3T3 cells. (a) NIH 3T3 fibroblasts expressing the UBE2L3-KRAS fusion lost normal fibroblast morphology. **(b)** Increased S phase of fusion expressing cells in contrast to the pDEST40 vector control. Polyclonal populations of NIH 3T3 cells transfected with either the pDEST40 vector or the UBE2L3-KRAS fusion were subjected to cell cycle analysis. Flow cytometry was carried out using propidium iodide staining.



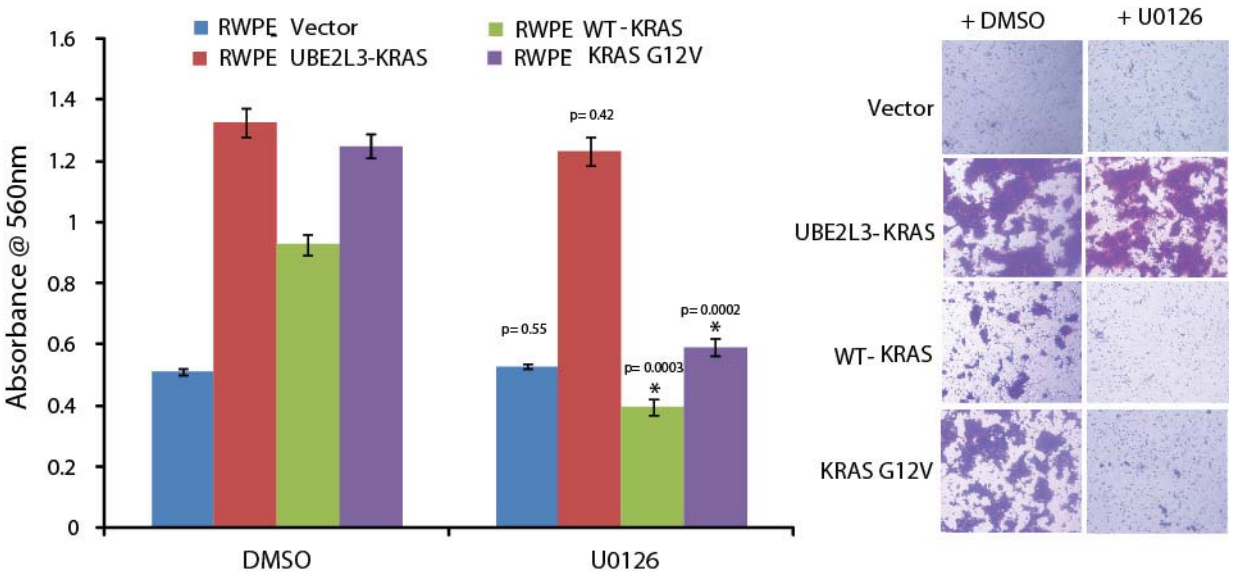
Supplementary Figure 10. Photographs and pathology of NIH 3T3 xenograft models. (a) The photographs of the mice bearing the NIH 3T3 xenograft tumors expressing UBE2L3-KRAS fusion (upper) and the pDEST40 vector (lower). **(b)** The pathology of NIH 3T3 xenograft tissues. Left panel, xenograft tissues excised from NIH 3T3 fusion expressing tumor bearing mice were stained using hematoxylin and eosin (HE). The tumors in the upper panel show poorly differentiated cells with increased nuclear size. A mass of cells obtained from only one mouse in the control group was also stained and the upper panel shows the fibroblast nature of the cells (lower). Right panel, Ki-67 immunohistochemical (IHC) staining of xenograft tissues showing 98% of tumor nuclei (upper) versus 17% of control tissue nuclei (lower) are positive for Ki-67.



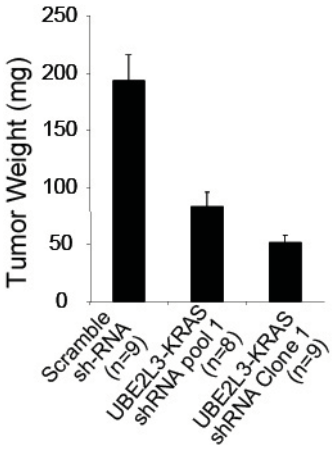
Supplementary Figure 11. Immunofluorescence study of stably transfected NIH 3T3 cells using monoclonal antibodies against KRAS (green channel), and a late endosome marker, LAMP1 (red channel). The colocalizing signals of KRAS and LAMP1 will be in yellow color. This image is based on mouse anti-KRAS monoclonal antibody (Invitrogen, Cat #: 415700), and rabbit anti-LAMP1 polyclonal antibody (Santa Cruz Biotechnology, Cat #: sc-5570). This result was reconfirmed using two additional antibodies: anti-KRAS mAb (Sigma-Aldrich, Cat#: WH0003845M1), and anti-RAS mAb (Millipore Cat #: 05-516).



Supplementary Figure 12. The UBE2L3-KRAS infected RWPE cells form transient tumors in mice. Stable monoclonal populations of RWPE cells expressing either the vector or UBE2L3-KRAS fusion gene were injected subcutaneously into nude mice. Tumor growth was assessed on day 6 as indicated.



Supplementary Figure 13. MEK inhibitor (U0126) has no effect on UBE2L3-KRAS fusion transcript mediated cell invasion. Stable RWPE cells over-expressing vector pDEST40, fusion transcript UBE2L3-KRAS, wild type WT-KRAS or oncogenic KRAS-G12V were seeded in the presence of DMSO or U0126 (20μM) for 48 hours in matrigel coated transwell inserts for invasion assay. Mean values obtained at 560nm absorbance from each condition were plotted on the y-axis.



Supplementary Figure 14. Knock-down of UBE2L3-KRAS fusion attenuates prostate tumor growth in a chicken embryo chorioallantoic membrane (CAM) model. Cells were inoculated on the CAM of 10-day-old chicken embryos. The chicken embryos were sacrificed on day 18 and the tumors formed on the upper CAM were collected and weighed.

Supplementary Tables

Supplementary Table 1. Oligonucleotide primers used for RT-PCR, qPCR and cloning.

Gene	Refseq.	Primer	Type	Bases	Exon	Sequence(5' to 3')	application
SOX5	NM_006940	SOX5_S1	Sense	456	Exon 3	ACAGAGTGGCGAGTCCTTGTCT	RT-PCR
SOX5	NM_006940	SOX5_S2	Sense	1324	Exon 10	TCACCCACATCACCCACCTCTC	RT-PCR
C14orf166	NM_016039	C14orf166_S1	Sense	143	Exon1	AAGTTGACGGCTCTCGACTACC	RT-PCR
C14orf166	NM_016039	C14orf166_S2	Sense	336	Exon3	GTCCTTTCAAGATTCAAGATCG	RT-PCR
UBE2L3	NM_003347	UBE2L3_S1	Sense	53	Exon 1	GGGAAGGAGCAGCACCAAATCC	RT-PCR
UBE2L3	NM_003347	UBE2L3_S2	Sense	76	Exon 1	AGATGGCGGCCAGCAGGAGGCT	RT-PCR
UBE2L3	NM_003347	UBE2L3_S3	Sense	316	Exon 3	ACGAAAAGGGGCAGGTCTGTCT	RT-PCR
UBE2L3	NM_003347	UBE2L3_q1	Sense	345	Exon 3	ATTAGTGCCGAAAACTGGAAGC	RT-PCR, qPCR
UBE2L3	NM_003347	UBE2L3_R1	Reverse	363	Exon 3	GGTCGGTTTTGGTTGCTGGCTT	RT-PCR
KRAS	NM_004985	KRAS_S1	Sense	204	Exon 2	TAGTTGGAGCTGGTGGCGTAGG	RT-PCR
KRAS	NM_004985	KRAS_R4	Reverse	204	Exon 2	CCTACGCCACCAGCTCCAACTA	RT-PCR
KRAS	NM_004985	KRAS_R3	Reverse	228	Exon 2	AGCTGTATCGTCAAGGCACTCT	RT-PCR
KRAS	NM_004985	KRAS_q2	Reverse	349	Exon 3	CTCCTCTTGACCTGCTGTGTCG	RT-PCR, qPCR
KRAS	NM_004985	KRAS_R7	Reverse	540	Exon 4	GTGTCTACTGTTCTAGAAGGCA	RT-PCR
KRAS	NM_004985	KRAS_R2	Reverse	1594	Exon 5	AGAGCAGTCTGACACAGGGAGA	RT-PCR
KRAS	NM_004985	KRAS_R1	Reverse	1893	Exon 5	GTCAGCAGGACCACCACAGAGT	RT-PCR
KRAS	NM_004985	KRAS_R6	Reverse	3313	Exon 5	ACTGGCATCTGGTAGGCACTCA	RT-PCR
GAPDH	NM_002046	GAPDH_S1	Sense	822	Exon 8	GTCAGTGGTGGACCTGACCT	RT-PCR
GAPDH	NM_002046	GAPDH_R1	Reverse	1014	Exon 8	TGAGCTTGACAAAGTGGTCG	RT-PCR
GAPDH	NM_002046	GAPDH_S2	Sense	556	Exon 7	TGCACCACCAACTGCTTAGC	qPCR
GAPDH	NM_002046	GAPDH_R2	Reverse	622	Exon 7	GGCATGGACTGTGGTCATGAG	qPCR
UBE2L3	NM_003347	N-FLAG	Sense	81	Exon 1	TCCACCATGGATTACAAGGATGACGACGATAA GGCGGCCAGCAGGAGGCTGATGAAGGAGCTTG	Cloning
UBE2L3	NM_003347	wild type	Sense	78	Exon 1	TCCACCATGGCGGCCAGCAGGAGGCTGATGAA GGAGCTTG	Cloning
KRAS4B	NM_004985	wild type 4B	Reverse	717	Exon 5	TTACATAATTACACACTTTGTCTTTGACTTCT	Cloning
KRAS4A	NM_033360	wild type 4A	Reverse	720	Exon 5	TTACATTATAATGCATTTTTTAATTTTCACAC	Cloning

Supplementary Table 2. The result of ABRA ranking analysis in a panel of 36 leukemia cell lines. Cell lines that do not harbor partially amplified genes are not shown.

Cell line	Gene	Chr	Gene position	Breakpoint position	Type	Level of amplification*	ConSig Score	Breakpoint curation#	Cancer Gene$
K-562	*ABL1*	chr9	132579089-132752883	132538227-132601560	3'amp	2.43	1.70		Yes
Kasumi-1	*ZBTB20*	chr3	115540207-116348817	115654416-115688377	3'amp	0.97	1.20		
697	*PBX1*	chr1	162795561-163082934	163015749-163027242	3'amp	0.88	1.18		Yes
U-937	*ZNF595*	chr4	43227-78099	34101-59713	3'amp	0.97	1.10		
U-937	*ZNF718*	chr4	43250-146491	34101-59713	3'amp	0.97	1.10		
ME-1	*LASS6*	chr2	169021081-169339398	169307675-169332507	3'amp	0.92	0.98		
SD1	*TRA@*	chr14	21159897-22090915	21480460-21483625	3'amp	0.82	0.80		
YT	*SMYD4*	chr17	1629603-1679844	1668363-1723617	3'amp	3.15	0.74		
NOMO-1	*AUTS2*	chr7	68702255-69895790	68965752-69022498	3'amp	1.04	0.73		
SKNO-1	*PPM1E*	chr17	54188231-54417319	54195586-54264925	3'amp	0.80	0.71		
Kasumi-1	*DCUN1D4*	chr4	52404033-52477760	48758236-52409268	3'amp	0.81	0.66		
UOCB1	*TRIT1*	chr1	40079315-40121764	40081505-40116019	3'amp	1.33	0.63		
NB4	*RASIP1*	chr19	53915654-53935782	53932963-53999660	3'amp	2.49	0.61		
Kasumi-1	*BOC*	chr3	114414065-114488996	114397446-114403958	3'amp	0.89	0.59		
Jurkat	*LSAMP*	chr3	117011832-117647068	117435489-117435958	3'amp	1.00	0.56		
UOCB1	*SSH2*	chr17	24977091-25281144	25114150-25145292	3'amp	1.43	0.56		
NB4	*CD72*	chr9	35599976-35608408	35605896-35672585	3'amp	1.80	0.50		
CMK	*TNFSF18*	chr1	171277074-171286679	171279427-171335867	3'amp	1.12	0.46		
NOMO-1	*GNE*	chr9	36204438-36248401	36227241-36296809	3'amp	1.16	0.45		
YT	*NMT1*	chr17	40494206-40541910	40524918-40529353	3'amp	1.08	0.43		
K-562	*GRID1*	chr10	87349292-88116230	87835186-87847802	3'amp	1.03	0.38		
BV173	*CRB1*	chr1	195504031-195714208	195630571-195665893	3'amp	0.91	0.34		
PL21	*KIF21A*	chr12	37973297-38123185	38082147-38177337	3'amp	1.08	0.32		
SKNO-1	*TMEM135*	chr11	86426713-86712220	86486132-86489947	3'amp	0.76	0.27		
Kasumi-1	*TMEM100*	chr17	51151989-51155141	51157187-51193501	3'amp	0.91	0.27		
NB4	*TMTC1*	chr12	29545024-29828959	29633231-29639626	3'amp	1.05	0.25		
CMK	*DEFB115*	chr20	29309128-29311096	28119554-29309964	3'amp	1.52	0.18		
BV173	*IGL@*	chr22	20710659-21595085	20987900-21049884	3'amp	1.15	0.16		
Kasumi-1	*C17orf57*	chr17	42756346-42873677	42768383-42771996	3'amp	0.80	0.13		
SKNO-1	*FOXD4L2*	chr9	69465527-69468635	67813967-68171592	3'amp	0.86	0.00		
MV4-11	*C1orf150*	chr1	245779072-245806482	245731777-245732788	3'amp	0.80	0.00		
K-562	*DGCR5*	chr22	17338027-17362141	17317513-17347582	3'amp	1.51	0.00		
BV173	*IGLV7-46*	chr22	21054162-21054455	20987900-21049884	3'amp	1.15		Not acceptable (5)	
CMK	*CHEK1*	chr11	125001547-125030847	125017829-125050645	3'amp	0.86		Not acceptable (6)	
CMK	*UBE4B*	chr1	10015630-10163884	10145592-10158626	3'amp	0.76		Not acceptable (6)	
CMK	*DNM3*	chr1	170077261-170648480	170339258-170384303	3'amp	1.21		Not acceptable (5)	
CMK	*ERI3*	chr1	44459329-44593526	44470260-44503655	3'amp	0.92		Not acceptable (5)	
Jurkat	*TERT*	chr5	1306282-1348159	1322006-1375087	3'amp	0.77		Not acceptable (6)	
Kasumi-1	*REST*	chr4	57468799-57493097	57443410-57460303	3'amp	1.67		Not acceptable (5)	
MV4-11	*LOC646479*	chr8	17333702-17373392	17367002-17367251	3'amp	1.34		Not acceptable (5)	
MV4-11	*CCDC25*	chr8	27646752-27686089	27660240-27663535	3'amp	1.20		Not acceptable (5)	
NOMO-1	*STARD13*	chr13	32575307-32757892	32941750-32950412	3'amp	0.84		Not acceptable (4)	
PL21	*FAM155A*	chr13	106618880-107317084	107054542-107088262	3'amp	2.21		Not acceptable (5)	
SD1	*TRAV15*	chr14	21488174-21488717	21480460-21483625	3'amp	0.82		Not acceptable (5)	
SKNO-1	*FNDC3B*	chr3	173312936-173601181	173467998-173474266	3'amp	1.09		Not acceptable (1)	
SUPB-15	*FAM49B*	chr8	130922898-131021182	130947000-131041167	3'amp	2.18		Not acceptable (5)	

TOM-1	*LOC729894*	chr15	20304046-20344434	20329239-20335459	3'amp	0.77	Not acceptable (6)
YT	*IRF4*	chr6	336760-356193	323970-340634	3'amp	0.89	Not acceptable (6)
YT	*SKAP1*	chr17	43565804-43862551	43592989-43599947	3'amp	0.93	Not acceptable (3)

[*] level of amplification shows the difference of relative quantification of DNA copy number data at the amplification breakpoints

[#] Situations when breakpoint is not acceptable

(1) Multiple intragenic breakpoints;

(2) The candidate is not the gene closest to the amplification breakpoint;

(3) The amplification starts from existing copy number increase and the breakpoint is not sharp;

(4) The breakpoint locates at the centromere or the end of the chromosome;

(5) The breakpoint is the result of a small deletion within an amplification;

(6) The breakpoint is found in a majority of samples.

[$] Cancer genes are defined by cancer gene census (http://www.sanger.ac.uk/genetics/CGP/Census/) [2]

Supplementary Table 3. The result of ABRA ranking analysis in a panel of 10 prostate cancer cell lines. Cell lines that do not harbor partially amplified genes are not shown.

Cell line	Gene	Chr	Gene position	Breakpoint position	Type	Level of amplification*	ConSig Score	Breakpoint curation#	Cancer Gene$
DU145.1	*KRAS*	chr12	25249446-25295121	25289615-25308034	3'amp	1.10	1.51		Yes
NCI660	*CREB5*	chr7	28112179-28638749	28602533-28619165	3'amp	1.05	1.23		
PC3	*ZNF605*	chr12	132108397-132143218	132153240-132181375	3'amp	0.95	1.02		
NCI660	*STRN3*	chr14	30432761-30565340	30459890-30475453	3'amp	0.79	0.88		
NCI660	*CBX2*	chr17	75366587-75375973	75345225-75368067	3'amp	0.96	0.87		
VCaP	*MYO16*	chr13	108046500-108658356	108470109-108485416	3'amp	1.48	0.84		
NCI660	*ANXA11*	chr10	81904859-81955308	81952099-81996286	3'amp	3.00	0.84		
PC3	*CHAF1A*	chr19	4353659-4394393	4343191-4350902	3'amp	0.86	0.80		
CA.hpv.10	*NEK7*	chr1	194933338-195020426	194997907-195020336	3'amp	1.29	0.78		
MDAPCa.2b	*POLB*	chr8	42315186-42348470	42308627-42321600	3'amp	1.06	0.73		
C4.2B	*TNFSF12*	chr17	7393098-7401930	7388929-7401533	3'amp	0.86	0.71		
DU145.1	*GPSM3*	chr6	32266521-32271278	32268003-32297149	3'amp	0.89	0.70		
PC3	*PARG*	chr10	50696332-51041337	50717613-50764988	3'amp	1.58	0.70		
PC3	*PTPN20A*	chr10	45970128-48447930	46371243-46396163	3'amp	0.76	0.69		
C4.2B	*CHFR*	chr12	132027287-132074534	132043499-132057649	3'amp	0.78	0.69		
NCI660	*POLQ*	chr3	122632963-122747519	122747877-122772579	3'amp	0.78	0.62		
NCI660	*MIS12*	chr17	5330970-5334852	5329458-5332851	3'amp	1.10	0.60		
PC3	*RINT1*	chr7	104766482-104802075	104742899-104771013	3'amp	0.89	0.59		
PC3	*DHX8*	chr17	38916859-38957206	38922343-38940540	3'amp	0.90	0.55		
C4.2B	*RHOG*	chr11	3804787-3818760	3823067-3843370	3'amp	0.78	0.55		
NCI660	*DULLARD*	chr17	7087882-7095983	7104658-7116744	3'amp	0.99	0.54		
NCI660	*KRT7*	chr12	50913220-50928976	50888386-50904894	3'amp	1.58	0.54		
PC3	*PHLDA3*	chr1	198166279-198169956	198168603-198184752	3'amp	1.87	0.53		
PC3	*HEATR4*	chr14	73014945-73095404	73082250-73092018	3'amp	1.22	0.51		
DU145.1	*CCDC130*	chr19	13719752-13735106	13663749-13726338	3'amp	1.10	0.44		
DU145.1	*PRSS36*	chr16	31057749-31068888	31065342-31081708	3'amp	0.88	0.43		
NCI660	*CLDN7*	chr17	7104183-7106513	7104658-7116744	3'amp	0.99	0.31		
NCI660	*CLDN7*	chr17	7104183-7106513	7104658-7116744	3'amp	0.99	0.31		
C4.2B	*TNFSF12-TNFSF13*	chr17	7393139-7405649	7388929-7401533	3'amp	0.86	0.29		
DU145.1	*SLC25A21*	chr14	36218828-36711616	36250273-36264559	3'amp	1.03	0.23		
PC3	*CCDC109A*	chr10	74121894-74317456	74212793-74231229	3'amp	2.10	0.22		
VCaP	*SIL1*	chr5	138310310-138561964	138352012-138377215	3'amp	1.23	0.21		
PC3	*PTPN20B*	chr10	45970128-48447930	46371243-46396163	3'amp	0.76	0.18		
NCI660	*TMLHE*	chrX	154283476-154406301	154405100-154429859	3'amp	1.29	0.11		
MDAPCa.2b	*CCDC36*	chr3	49210864-49270159	49200389-49211507	3'amp	0.76	0.00		
VCaP	*C6orf106*	chr6	34663049-34772603	34757380-34808070	3'amp	1.91	0.00		
X22RV1	*DTWD2*	chr5	118203134-118352139	118249938-118267673	3'amp	0.84	0.00		
C4.2B	*UBE2T*	chr1	199032442-199042734	199032584-199065829	3'amp	1.12		Not acceptable (5)	
C4.2B	*PPP1R12B*	chr1	199049492-199289354	199194296-199208949	3'amp	1.01		Not acceptable (5)	
C4.2B	*SENP3*	chr17	7406042-7416009	7388929-7401533	3'amp	0.86		Not acceptable (2)	
C4.2B	*TNFSF13*	chr17	7402339-7405641	7388929-7401533	3'amp	0.86		Not acceptable (2)	
DU145.1	*PBX2*	chr6	32260495-32265941	32268003-32297149	3'amp	0.89		Not acceptable (2)	
DU145.1	*AGER*	chr6	32256723-32260001	32268003-32297149	3'amp	0.89		Not acceptable (2)	
NCI660	*HM13*	chr20	29565901-29621029	29557170-29575050	3'amp	0.91		Not acceptable (6)	
NCI660	*LOC541473*	chr7	74665887-74669359	73431763-73452181	3'amp	0.77		Not acceptable (5)	
NCI660	*STAG3L2*	chr7	71913147-73751326	73431763-73452181	3'amp	0.77		Not acceptable (3)	
PC3	*KLHL17*	chr1	936109-941162	919074-931088	3'amp	0.80		Not acceptable (5)	

X22RV1	*UBE2T*	chr1	199032442-199042734 199032584-199065829 3'amp	0.85	Not acceptable (5)
X22RV1	*PPP1R12B*	chr1	199049492-199289354 199194296-199208949 3'amp	0.77	Not acceptable (5)

*, #, $ See the notes for Supplementary Table 1.

Supplementary Table 4. Matching the amplification level of 5' amplified genes with *ABL1* and *KRAS* on K-562 and DU145 respectively nominates their candidate 5' partners. 3' genes seeding the breakpoint assembling analysis are highlighted by bold.

Cell line	Gene	Chr	Gene position	Breakpoint position	Type	Level of amplification*	Breakpoint curation**
K-562	*BCR*	chr22	21852552-21990224	21953284-21965924	5'amp	2.89	
K-562	*FBXW4P1*	chr22	21934791-21937180	21953284-21965924	5'amp	2.89	Not acceptable (2)
K-562	*GPC5*	chr13	90848930-92317491	91268794-91293729	5'amp	2.89	Not acceptable (1)
K-562	*NUP214*	chr9	132990802-133098912	133105930-133147163	5'amp	2.57	
K-562	***ABL1***	**chr9**	**132579089-132752883**	**132538227-132601560**	**3'amp**	**2.43**	
K-562	*CAMP*	chr3	48239866-48241979	48235972-48294562	5'amp	1.22	level not matched
DU145.1	*SOX5*	chr12	23576498-24606647	23557809-23576870	5'amp	1.14	
DU145.1	***KRAS***	**chr12**	**25249446-25295121**	**25289615-25308034**	**3'amp**	**1.10**	
DU145.1	*C14orf166*	chr14	51525942-51667191	51530172-51541520	5'amp	1.04	
DU145.1	*PPAP2C*	chr19	232045-242435	64418-232080	5'amp	1.03	Not acceptable (4)
DU145.1	*RNF5*	chr6	32254149-32256545	32268003-32297149	5'amp	0.89	Not acceptable (2)
DU145.1	*UBE2L3*	chr22	20246510-20302877	20289615-20302254	5'amp	0.89	
DU145.1	*MYST1*	chr16	31036487-31050206	31065342-31081708	5'amp	0.88	Not acceptable (2)
DU145.1	*BTRC*	chr10	103103814-103307058	103107408-103128976	5'amp	0.86	Not acceptable (3)
DU145.2	*SOX5*	chr12	23576498-24606647	23480427-23557809	5'amp	0.99	
DU145.2	*CCDC116*	chr22	20311639-20316169	20313694-20322692	5'amp	0.97	Not acceptable (2)
DU145.2	*SDF2L1*	chr22	20321095-20323141	20313694-20322692	5'amp	0.97	Not acceptable (2)
DU145.2	*UBE2L3*	chr22	20246510-20302877	20313694-20322692	5'amp	0.97	
DU145.2	***KRAS***	**chr12**	**25249446-25295121**	**25289615-25308034**	**3'amp**	**0.96**	
DU145.2	*C14orf166*	chr14	51525942-51667191	51530172-51541520	5'amp	0.76	

Supplementary References

1. Olshen, A.B., Venkatraman, E.S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557-572 (2004).
2. Futreal, P.A.*, et al.* A census of human cancer genes. *Nat Rev Cancer* **4**, 177-183 (2004).
3. Wang, X.-S.*, et al.* Integrative Analyses Reveal the Functional and Genetic Associations of Gene Fusions in Cancer. in *submitted* (2009).
4. Korenchuk, S.*, et al.* VCaP, a cell-based model system of human prostate cancer. *In Vivo* **15**, 163-168 (2001).
5. Maher, C.A.*, et al.* Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc Natl Acad Sci U S A* (2009).
6. Rubin, M.A.*, et al.* Overexpression, amplification, and androgen regulation of TPD52 in prostate cancer. *Cancer Res* **64**, 3814-3822 (2004).
7. Garraway, L.A.*, et al.* Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117-122 (2005).
8. Witkiewicz, A.K.*, et al.* Alpha-methylacyl-CoA racemase protein expression is associated with the degree of differentiation in breast cancer using quantitative image analysis. *Cancer Epidemiol Biomarkers Prev* **14**, 1418-1423 (2005).