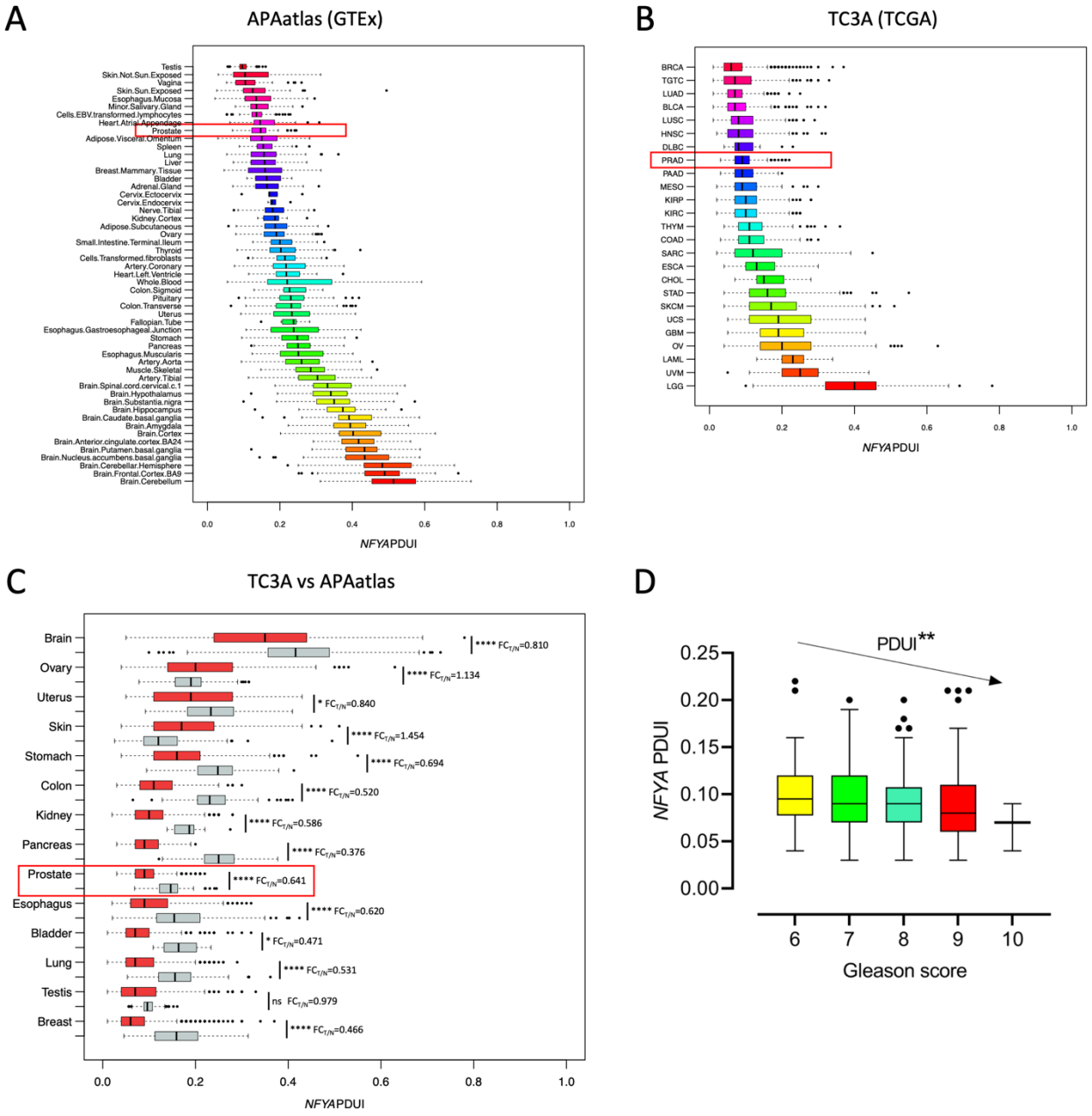
SUPPLEMENTARY INFORMATION

Supplementary Figures

Supplementary Methods

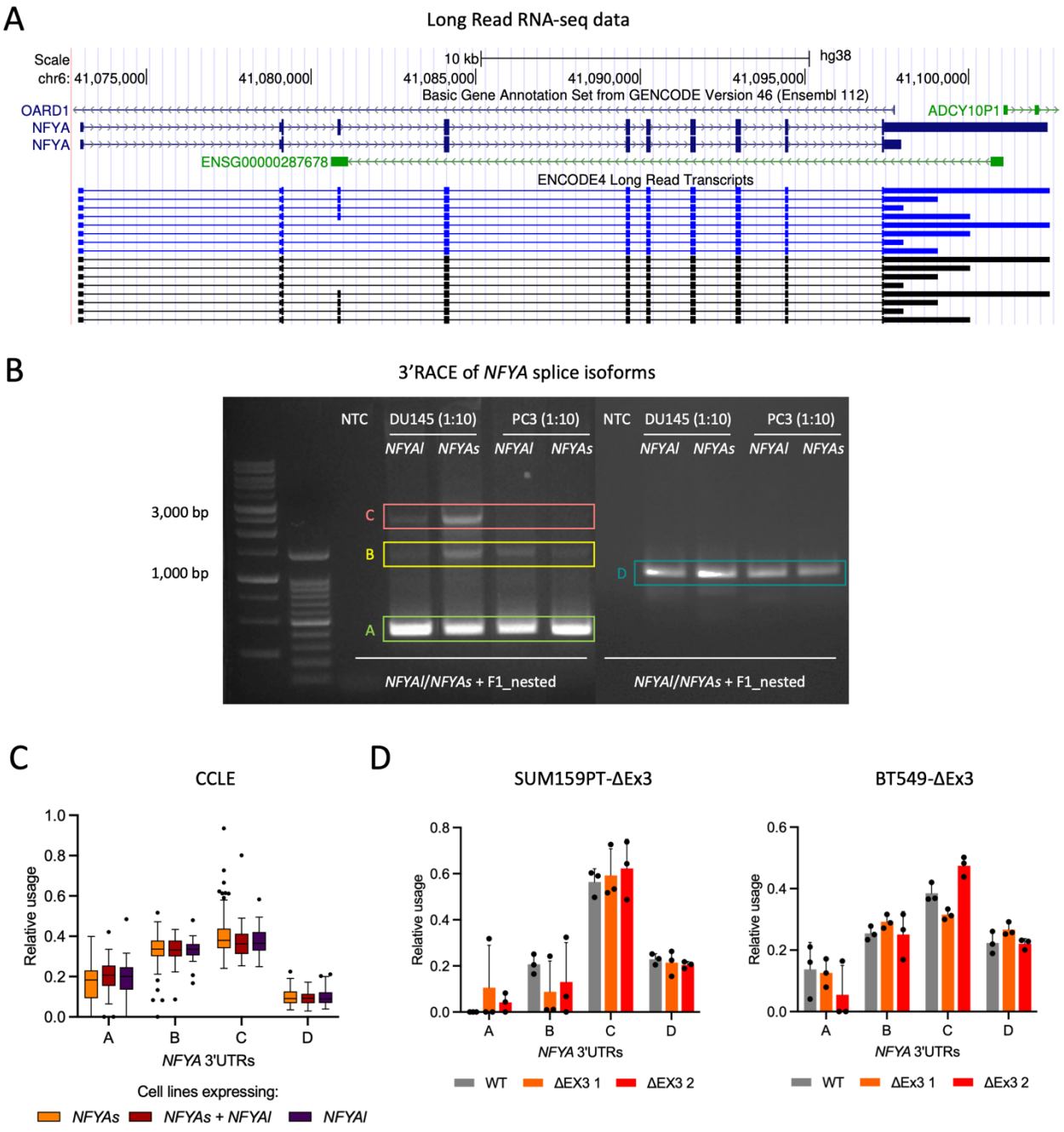
Supplementary Tables

SUPPLEMENTARY FIGURES



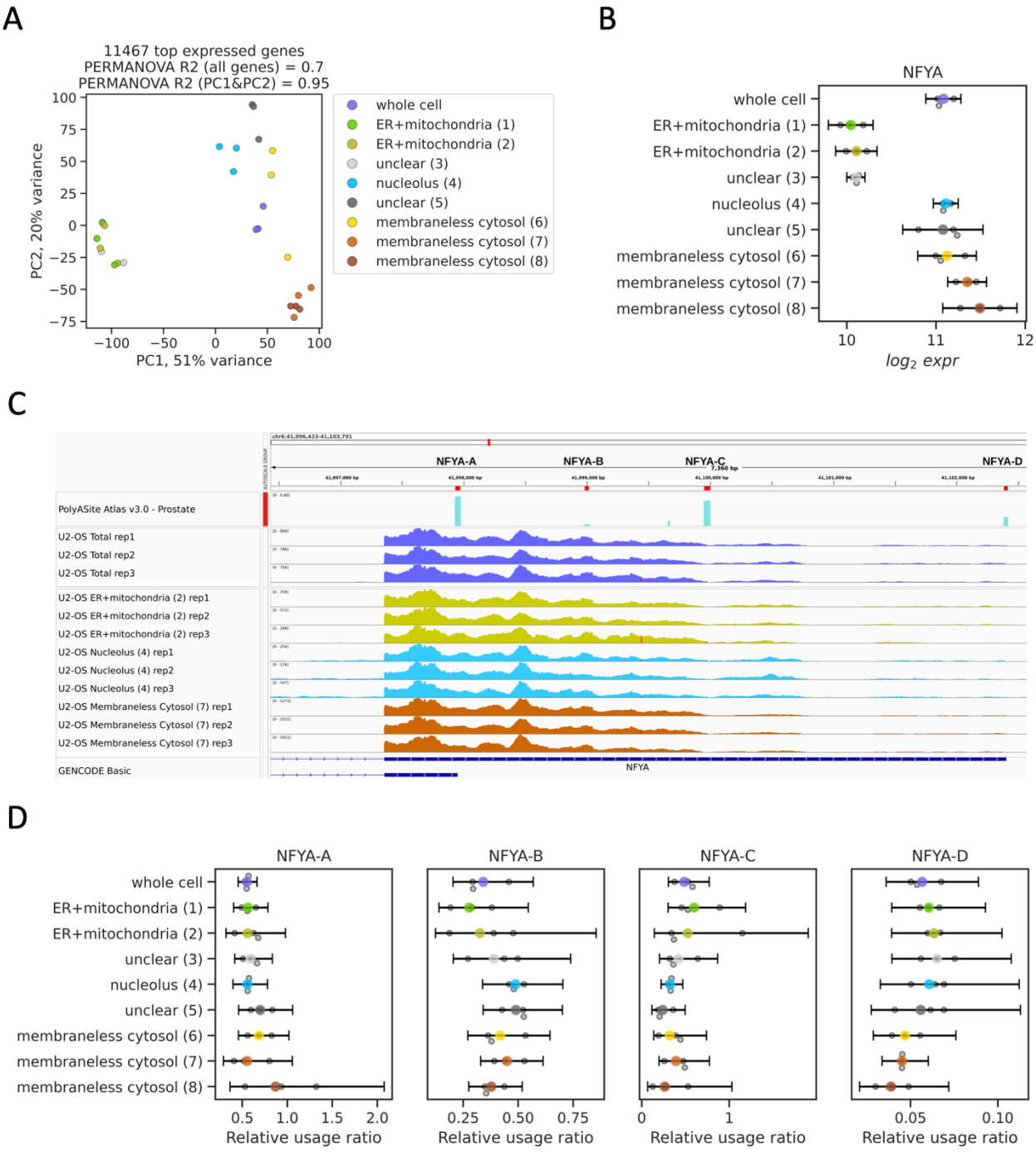
Supplementary Figure S1: *NFYA* 3'UTR usage in normal and tumor tissues. (**A**) Boxplot showing the Percentage of Distal polyA site Usage index (PDUI) of *NFYA* in GTEx normal tissues as from APAatlas and (**B**) in TCGA tumor tissues as from TC3A. PDUI varies from 0 to 1: when it tends to 1, the distal 3'UTR is more used, denoting a lengthening event, while when it tends to 0, the proximal 3'UTR is more used, denoting a shortening event. (**C**) Boxplots showing the comparison between distributions of PDUI values across normal

and tumor tissues as form APAatlas and TC3A, respectively. The fold-change of the average PDUI values in tumor (red boxes) and normal (grey boxes) tissues ($FC_{T/N}$) calculated for *NFYA* and the two-tailed unpaired *t*-test *p*-values are reported. (**D**) Boxplot showing *NFYA* PDUI values in PCa samples from TC3A, split according to the tumor grade (i.e. Gleason score). The Jonckheere-Terpstra test *p*-values are reported on the graphs.



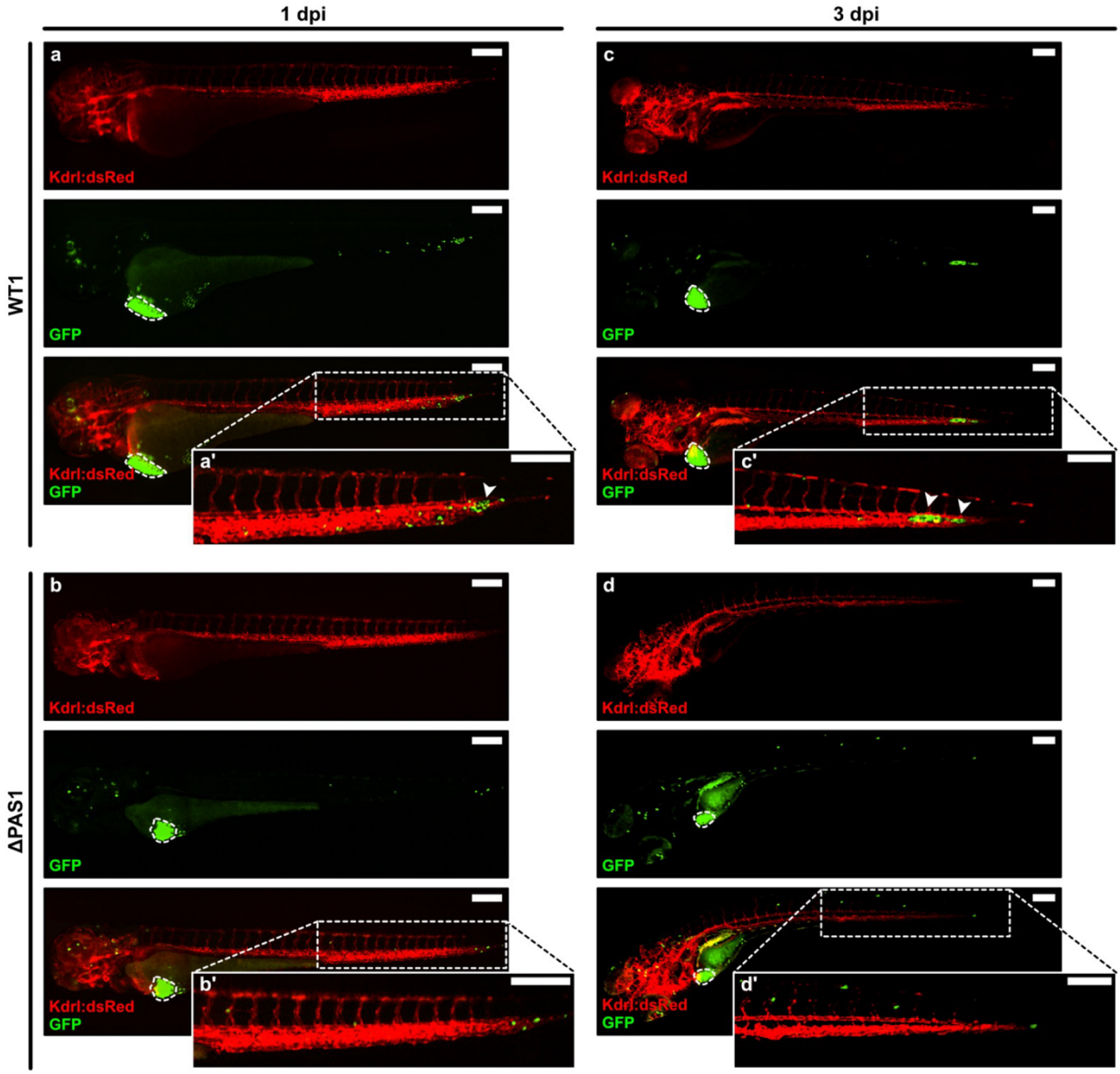
Supplementary Figure S2: Association between *NFYA* splicing and APA isoforms. (**A**) *NFYA* locus at chr6 (p21.1) as from UCSC Genome Browser (GRCh38/hg38 assembly) showing track for ENCODE4 Long Read Transcripts aligned to GENECODE V46 annotation. (**B**) Electrophoretic gel of 3'RACE experiments using *NFYA* splicing isoform-specific primers (NFYA_long_F or NFYA_short_F) for the first PCR round and F1_nested or F4_nested primers for the second PCR round starting from a PCR dilution of 1:10. NTC = no template control. The bands corresponding to the different *NFYA* 3'UTRs are highlighted in the figure: 3'UTR-A, green, 3'UTR-

B, yellow, 3'UTR-C, red, 3'UTR-D, blue. (**C**) Boxplot showing the relative usage of *NFYA* 3'UTRs, quantified on our new annotation as described in Supplementary Methods, in 233 cancer cell lines (CCLE data) stratified according to *NFYAl* and *NFYAs* expression. Cell lines mainly expressing *NFYAs* defined as *NFYAl*/*NFYAs* < 0.5; cells mainly expressing *NFYAl* defined as *NFYAl*/*NFYAs* > 1.5; cells expressing both *NFYAs* + *NFYAl* defined as 0.5 <= *NFYAl*/*NFYAs* <= 1.5. (**D**) Barplots showing the relative usage of *NFYA* 3'UTRs in SUM59PT (*left*) and BT549 (*right*) breast cancer cell lines deleted for *NFYA* exon 3 (ΔEx3) (GSE208088). For each cell line, two deleted clones (ΔEx3) and one non-deleted clone (WT) were shown.



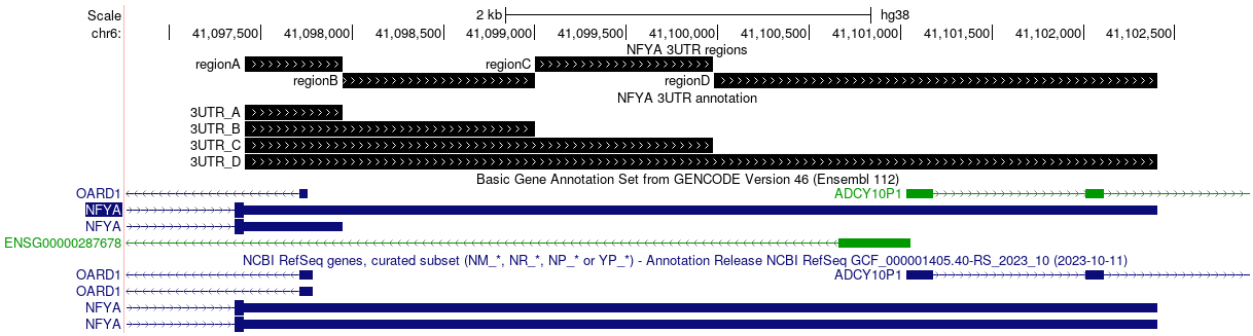
Supplementary Figure S3. Analysis of subcellular RNA localization from the Villanueva *et al.* study. (**A**) PCA analysis of pySanity-normalized gene expression values. Axis labels indicate the proportion of total variance explained by each principal component. Permutational multivariate analysis of variance (PERMANOVA) across all 11,467 input genes yielded an adjusted R^2 of 0.70, while analysis restricted to the first two principal components yielded an adjusted R^2 of 0.95. Together, these results demonstrate strong reproducibility

among biological replicates and clear separation between distinct subcellular fractions. As expected, the whole-cell lysate samples cluster centrally relative to the isolated fractions. (**B**) *NFYA* pySanity-normalized gene expression levels. Error bars represent 95% confidence intervals, Bonferroni-adjusted for multiple comparisons across all displayed conditions. *NFYA* expression is significantly lower in fractions (1)-(3) compared to other fractions and whole cell lysate, as expected for mRNA of a nuclear transcription factor-encoding gene. In contrast, ER-proximal transcriptome preferentially codes for proteins that localize to the ER, Golgi, vesicles, and plasma membrane [1]. **(C)** Integrative genome browser (IGV [2]) screenshot of *NFYA* terminal exon. Displayed raw coverage was automatically calculated by IGV from input .bam files of selected RNA-seq samples. At the top, PASs corresponding to *NFYA* polyadenylation isoforms A, B, C, D are indicated. The average expression (reads-per-million, RPM) of PASs in healthy human prostate tissue from PolyASite v3.0 is also displayed. (**D**) Analysis of relative *NFYA* 3'UTR isoform usage across subcellular fractions. Error bars represent 95% confidence intervals, Bonferroni-adjusted for multiple comparisons across all displayed conditions. Relative usage of a particular *NFYA* isoform is calculated relative to the aggregated expression of other isoforms (see Supplementary Methods), hence the values are not limited by 0-1 range.

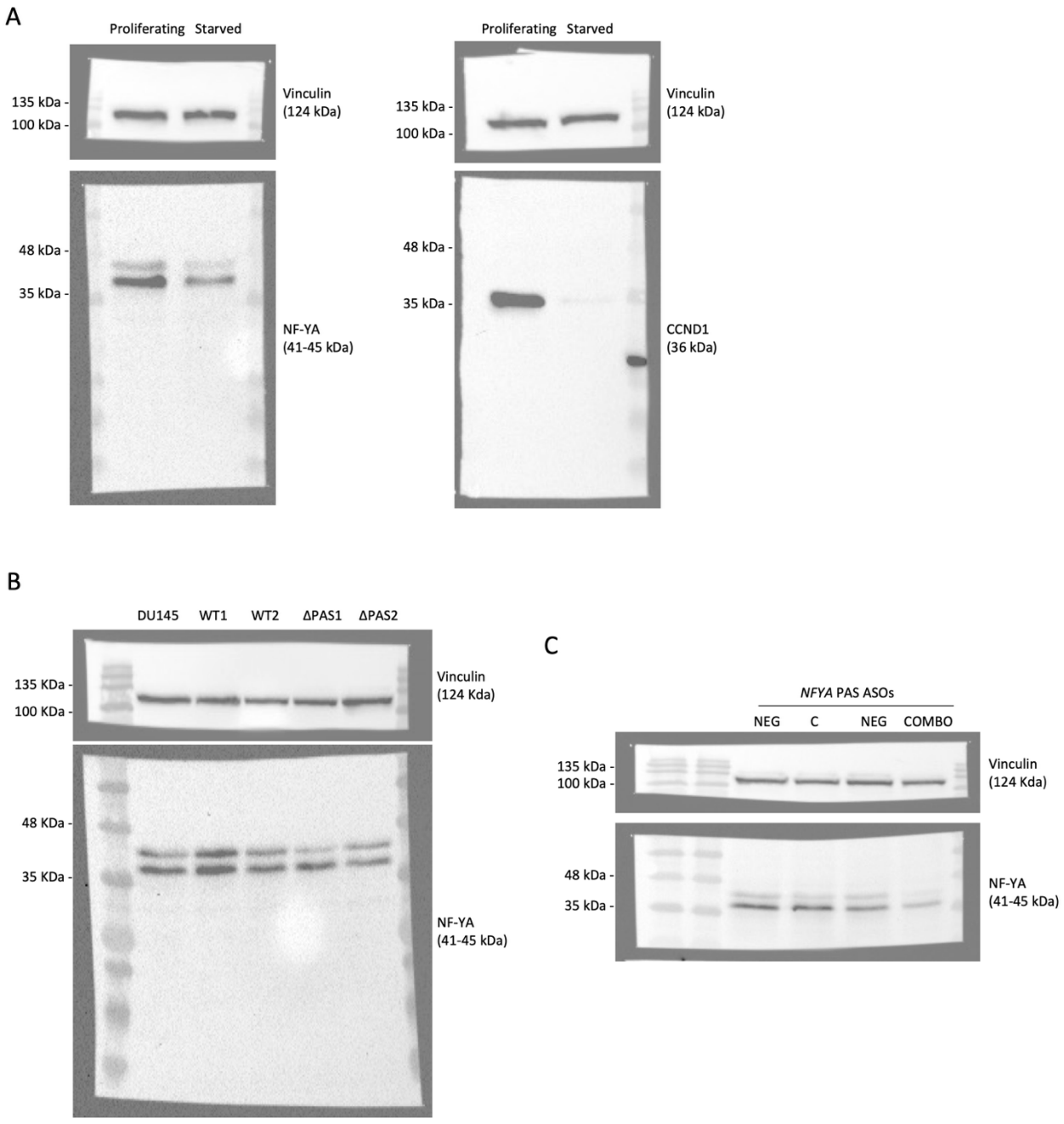


Supplementary Figure S4: APA manipulation through PAS deletion of *NFYA* 3'UTR-C impairs metastatic dissemination of PCa cells in zebrafish xenografts. Representative stereomicroscope images of live xenografts at 1 (a, b) and 3 (c, d) days post injection (dpi), showing engraftment and dissemination of either WT1 (a, c) or ΔPAS1 cells (b, d). Each image presents separated and merged views of the red and green channels, corresponding to Kdrl:dsRed and GFP signals, respectively. Dashed boxes outline the caudal hematopoietic tissue (CHT), which is magnified in panels a'–d'. Contrast enhancement was applied in magnifications to highlight cancer cell localization in the CHT. Both WT1 and ΔPAS1 cells formed primary

tumors (PTs) at the injection sites (outlined by dashed curves). WT1 cells formed metastases (clusters of ≥5 extravasated cells) as early as 1 dpi (a', c'; arrowheads), whereas ΔPAS1 cells were predominantly detected as single circulating cells at both 1 dpi (b') and 3 dpi (d'). Scale bars: 200 μm.



Supplementary Figure S5: Scheme of *NFYA* 3'UTR unique regions used for usage assessment. *NFYA* locus at chr6 (p21.1) as from UCSC Genome Browser (GRCh38/hg38 assembly). From the top to the bottom: Unique regions of *NFYA* 3'UTRs used as target windows to assess each single 3'UTR usage from RNA sequencing data; *NFYA* 3'UTR reconstructed annotation; GENCODE annotation (version 46) of *NFYA* 3'UTRs; NCBI RefSeq (release 229) annotation of *NFYA* 3'UTRs.



Supplementary Figure S6: Uncropped western blots. Uncropped western blots related to Figure 2H **(A)**, Figure 4D **(B)** and Figure 6C, H **(C)**.

SUPPLEMENTARY METHODS

Analysis of 3'UTR usage in publicly available databases. As for *NFYA* 3'UTR usage in publicly available datasets, we investigated APAatlas [3] for GTEx [4] normal tissues, and TC3A [5] for TCGA [6] tumor tissues. They refer to the NCBI RefSeq annotation (GRCh37/hg19 genome), which annotates transcripts with the longest possible 3'UTR, and exploit DaPars algorithm for the *de novo* identification of potential APA events based on localized changes in density of 3'UTR RNA-sequencing reads [7]. These report the Percentage of Distal polyA site Usage Index (PDUI) to express 3'UTR usage genome wide. Therefore, PDUI is a ratio of the usage of the distal (canonical) 3'UTR over the first detected proximal 3'UTR. It varies from 0 to 1: when tends to 1, the distal 3'UTR is more used, denoting a lengthening event, while when it tends to 0, the proximal 3'UTR is more used, denoting a shortening event.

Data download and processing. Transcriptomic data of a panel of 233 cancer cell lines from the Cancer Cell Line Encyclopedia (CCLE) [8](hereafter called CCLE data), for which proteomic data are also available, were downloaded from DepMap portal [9]. Moreover, we used SUM159PT and BT-549 breast cancer cell lines (GSE208088) [10], as well as tumor and normal prostate cells (GSE25183). FASTQ files of these paired-end RNA-sequencing data were downloaded using the fastq-dump utility of the SRA toolkit 2.9.4 version (https://sra-explorer.info/). FASTQ files containing raw sequencing reads were aligned to NCBI RefSeq (GRCh38/hg38 genome) using STAR and gene expression levels (expressed as transcript per million, TPM) and raw counts were computed using RSEM-1.3.1 software [11] using NCBI RefSeq (GRCh38/hg38) as a reference.

Transcriptomic data of two controlled datasets from The database of Genotypes and Phenotypes (dbGaP) [12, 13] were downloaded following the procedure indicated in the NCI Genomics Data Commons (GDC) Portal [14], with prior authorization to data access for this study.

In particular, the paired-end RNA-Seq FASTQ raw data files from TCGA-PRAD project [15] were downloaded from the GDC portal with the gdc-client tool (version 2.3) [16]. Read coverage for representative TCGA-PRAD samples reported in Figure 2D was made available via Recount3 resource [17–19]

A subselection of paired-end RNA-Seq FASTQ raw data files from the Integrative Clinical Sequencing Analysis of Metastatic Castration Resistant Prostate Cancer Reveals a High Frequency of Clinical Actionability project from the Stand Up To Cancer East Coast Prostate Cancer Research Group (hereafter called SU2C data) [20] were downloaded from the dbGaP portal with the fastq-dump tool from the SRA Toolkit (release 3.2.1) [21]. More in detail, only the 43 samples with the "LibrarySelection" feature set as "POLYA" were selected.

Reads were aligned to the reference NCBI RefSeq genome (GRCh38/hg38) with STAR (version 2.7.10a) [22].

***NFYA* 3'UTR annotation.** Alignment files of the selected data were sorted and indexed using "samtools sort" and "samtools index". Bedgraph files were created starting from sorted alignment files using "bedtools genomecov -ibam" command and filtered out for *NFYA* genomic coordinates using "bedtools intersect" command. *NFYA* bedgraph files were used as input for APAtrap tool [23], which predicts the presence of potential alternative 3'UTRs from bulk RNA-sequencing data. This analysis was performed on CCLE data using default parameters and results were visualized on the UCSC Genome Browser [24] (GRCh38/hg38 genome). Bigwig files containing the read coverage were generated using "bamCoverage" command starting from sorted alignment files for DU145 and PC3 data as from CCLE. Bigwig files were loaded and visualized on the UCSC Genome Browser (GRCh38/hg38 genome).

Displayed raw coverage for the SU2C mCRPC samples was automatically calculated by UCSC Genome Browser from input .bam files, restricted to chr6 with samtools view.

Bulk 3'end sequencing data were retrieved from polyA site 2.0 [25] and polyA DB V3.2 [26], which annotate *de novo* the possible alternative APA sites and report a score for their usage. Single cell RNA-sequencing data were also interrogated as from polyA site 3.0 [27]. We considered annotated APA sites only on the positive strand concordantly with *NFYA* annotation. Data were visualized on the UCSC Genome Browser, specifically

polyA site 2.0 and 3.0 data on GRCh38/hg38 genome, while polyA DB V3 data were translated to GRCh38/hg38 genome.

***NFYA* 3'UTR usage.** Starting from sorted and indexed alignment files of the selected data from CCLE, *NFYA* 3'UTR usage was evaluated considering the actual annotation with the four 3'UTRs and using "samtools view -c -f 16" command to count the reverse reads falling in each 3'UTR regions (A, B, C, and D) as depicted in Supplementary Figure S5. The number of reads of each region was normalized to the length in base pairs (bp) and subtracted according to the following scheme: A = region A – region B; B = region B – region C; C = region C – region D; D = region D, with A, B, C, and D being the absolute usage of each 3'UTR. The usage was finally expressed as relative usage normalized to the total usage of all 3'UTRs.

From alignment files of the selected data from TCGA-PRAD and SU2C libraries, properly paired reads with MAPQ greater than 20 and primary alignment (samtools view -f 0x2 -q 20 -F 0x904) were extracted. To then compute 3'UTR usage, we assigned them uniquely to each segment by sequentially removing reads overlapping with downstream 3'UTR regions as well, identified by read ID intersection. The number of reads for each segment was then normalized according to the 3'UTR region length in base pairs.

To analyze the association between *NFYA* splicing and APA isoforms, CCLE cancer cell lines were stratified according to the ratio of *NFYAl* and *NFYAs* expression: cells mainly expressing *NFYAs* = *NFYAl*/*NFYAs* < 0.5; cells mainly expressing *NFYAl* = *NFYAl*/*NFYAs* > 1.5; cells expressing both = 0.5 <= *NFYAl*/*NFYAs* <= 1.5.

Data from the CCLE, TCGA-PRAD and SU2C datasets were also split according to the 3'UTR usage expressed as the ratio of the usage of proximal 3'UTRs (A+B) over the usage of distal 3'UTRs (C + D); high ratio means that shorter 3'UTRs are more used, while low ratio means that distal 3'UTRs are more used. CCLE data were also split according to NF-YA protein levels, as from DepMap portal, and proliferation score, calculated using single sample GSEA analysis [28]. To perform differential analysis, we selected the top 20 cell lines and the bottom 20 cell lines according to these features and we evaluated 3'UTR usage, NF-YA protein level, and proliferation score.

Cell line cultures. Established human cancer cell lines purchased from American Type Culture Collection (ATCC, Rockville, MD, USA) were grown in standard conditions and checked for mycoplasma contamination using N-GRADE Mycoplasma PCR Reagent set (Euroclone S.p.A., Pero, Italy) and authenticated by STR profiling (Eurofins MWG-Biotech, Ebersberg, Germany). DU145 is a cell line with epithelial morphology that was isolated from the brain of a 69-year-old, white, male with prostate cancer. PC3 is a cell line initiated from a bone metastasis of a grade IV prostatic adenocarcinoma from a 62-year-old, white, male. Both cell lines were cultured in RPMI-1640 (Euroclone S.p.A., Pero, Italy) supplemented with 10% fetal bovine serum (FBS) (Euroclone S.p.A., Pero, Italy), 2 mM L-Glutamine (Euroclone S.p.A., Pero, Italy) and 25mM Hepes (Euroclone S.p.A., Pero, Italy), at 37 °C in a 5% CO_2 atmosphere. T6B-transduced cells were cultured in RPMI-1640 with Tetracycline Negative FBS (Euroclone S.p.A., Pero, Italy), 2 mM L-Glutamine, 25mM Hepes, and 125 µg/mL Hygromycin-B (Thermo Fisher Scientific Inc., Waltham, MA, USA). Phoenix Amphotropic (Phoenix-AMPHO) is a retrovirus producer cell line with epithelial morphology, generated from HEK293T cell line. Cells were maintained in DMEM (Euroclone S.p.A., Pero, Italy) supplemented with 10% FBS and 1% L-Glutamine, at 37 °C in a humidified atmosphere containing 5% CO_2. Other cell pellets were used: LNCaP, 22RV1 (prostate cancer), PNT1A, RWPE1 (prostate immortalized epithelial cells), WPMY-1 (prostate immortalized fibroblasts), and PrEC (prostate normal epithelial cells); we also used commercial RNA from bulk normal prostate tissue.

Starvation. 2×10^6 PC3 cells were seeded in T75 flask in 10 mL complete medium per sample. The day after, the medium was replaced with a medium without the serum (0.5% FBS) and L-Glutamine (0% L-Glut) for 24 hours of starvation as done in White *et al*. [29]. Proliferating cells grown with complete medium were considered as control and Cyclin D1 (CCND1) expression was measured as a proliferation marker [30].

RNA extraction, reverse transcription, and qRT-PCR. Total RNA was extracted form cellular pellets using the miRNeasy Mini Kit (QIAGEN, Hilden, Germany), including DNAse I digestion (QIAGEN, Hilden, Germany), according to manufacturer's instructions. RNA yield, A260/A280, A260/A230 ratios were monitored with Nabi-UV/VIS Nano Spectrophotometer MicroDigital (TWIN HELIX SRL, Rho, Italy). The ratio of A260/A280 and

A260/A230 in the range 1.8-2.0 were considered acceptable. RNA was reverse transcribed into cDNA using the M-MLV Reverse Transcriptase kit (GeneSpin Srl, Milan, Italy). Gene expression level was measured through qRT-PCR using SsoAdvancedTM Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA) and Bio-Rad CFX Connect™ RT-PCR instrument (Bio-Rad, Hercules, CA, USA). Gene expression level was assessed using the ΔΔCt method and expressed as relative quantity (rq). *RPS20* was used as endogenous gene; instead, *NFYA* was used for the normalization of the expression of *NFYA* 3'UTRs. All the primers were designed with Primer3Plus software [31] and their sequences are reported in Supplementary Table S1. As for *NFYA* 3'UTRs, we designed four different primer pairs that target the different regions of *NFYA* 3'UTRs: PrA targets 3'UTR-A region, PrB targets 3'UTR-B region, PrC targets 3'UTR-C region, and PrD targets 3'UTR-D region. It is important to note that these primers are not 3'UTR isoform-specific, except for PrD: indeed, PrA detects all the 3'UTRs; PrB detects 3'UTR-B, -C, and -D; PrC detects 3'UTR-C and -D; PrD detects only 3'UTR-D.

Protein extraction and western blot. For extraction of total proteins, cellular pellets were lysed in RIPA buffer (10mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 0.1% SDS, 0.1% Deoxycholic acid, 1% Triton X- 100, with the addition of 0.1% PIC (Roche by Sigma-Aldrich, Saint Louis, MI, USA) and 0.1% PMSF (Sigma-Aldrich, Saint Louis, MI, USA) protease inhibitors) for 30 min on ice and the supernatant (i.e., proteins) was boiled at 95 °C for 5 min. Proteins were loaded onto a 4-12% SDS-polyacrylamide gels (stacking gel: 125 mM Tris HCl pH= 6.8, 4% acrylamide:bis-acrylamide solution 37.5:1 (Euroclone S.p.A., Pero, Italy), 0.1% SDS, 0.1% APS, 0.01% Temed; separating gel: 375 mM Tris HCl pH= 8.8, 12% acrylamide:bis-acrylamide solution 37.5:1 (Euroclone S.p.A., Pero, Italy), 0.1% SDS, 0.1% APS, 0.01% Temed). Electrophoretic run was carried out at constant voltage (80-110V) in Running Buffer (25 mM Tris-Base, 190 mM Glycine, 0.1% SDS). Samples were transferred to a nitrocellulose membrane at constant current (0.35A) at 4°C for 2 hours in Transfer Buffer (25 mM Tris-Base, 190 mM Glycine, 20% methanol), using BioRad Mini-PROTEAN® Tetra Cell 4-Gel System. Nitrocellulose membranes were blocked for 1 hour shaking incubation in 5% BSA solution in TBS-T (50 mM Tris-Base, 150 mM NaCl, 0.1% Tween 20) at room temperature. Membranes were incubated with the primary antibody overnight at 4°C in continuous shaking. Primary antibodies were prepared in 5% BSA solution in TBS-T

according to recommended dilution as reported in Supplementary Table S3. Afterwards, membranes were washed three times in TBS-T for 10 minutes and incubated with the secondary antibody (anti-Rabbit or anti-Mouse) for 1 hour at room temperature in continuous shaking. Peroxidase-conjugated secondary antibodies (Sigma-Aldrich, Saint Louis, MI, USA) were prepared according to recommended dilution (1:10,000 in 5% BSA solution in TBS-T). Membranes were washed three times in TBS-T for 10 minutes with shaking at room temperature and briefly incubated in peroxidase solution (1:1 of Peroxidase and Enhancer Solution purchased by GeneSpin Srl, Milan, Italy). Immunoreactive bands were visualized using Bio-Rad ChemiDoc MP, with ImageLab software. Vinculin was used as housekeeping protein for normalization and quantification of bands intensity using ImageJ software. Uncropped western blots are reported as Supplementary Figure S6.

Cloning of *NFYA* 3'UTRs. pGL4.13-C with the chimeric luciferase with 3'UTR-C was obtained by digesting the original vector and PCR fragments of the 3'UTR-C with FseI restriction enzyme (New England Biolabs, Ipswich, UK). PCR fragments of 3'UTR-C were produced using Q5® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, UK) and 3'UTR-C-specific primers with the addition of FseI restriction sites. Ligation of digested vector and PCR amplicons was performed using T4 DNA Ligase (New England Biolabs, Ipswich, UK). Ligation reaction was used to transform STBL3 chemically competent E. coli cells (GeneSpin Srl, Milan, Italy), according to the manufacturer's instructions. pGL4.13-A and pGL4.13-B with the chimeric luciferase with 3'UTR-A and B were obtained starting from pGL4.13-C removing the regions 3'UTR A-C and B-C, respectively, using Q5® Site-Directed Mutagenesis Kit (New England Biolabs, Ipswich, UK). Finally, pGL4.13-D was obtained by cloning the entire 3'UTR-D fragment downstream of the *luc2* gene using NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs, Ipswich, UK) and 10-beta Competent E. coli (High Efficiency) (New England Biolabs, Ipswich, UK), according to manufacturer's instructions. The different grown bacterial colonies were pre-inoculated, expanded, and purified to extract the different vectors using QIAprep Spin Miniprep Kit (QIAGEN, Hilden, Germany). Obtained vectors from different colonies were screened by performing FseI digestion and verified through sequencing by Eurofins MWG-Biotech (Ebersberg, Germany). Correct vectors

were amplified starting from bacterial pre-inoculation using PureLinkTM HiPure Plasmid Midiprep Kit (Thermo Fisher Scientific Inc., Waltham, MA, USA).

Cell transfection. Luciferase reporter vectors were transiently transfected using Lipofectamine™ 2000 (Thermo Fisher Scientific Inc., Waltham, MA, USA) according to manufacturer's instructions. 1 x 10^6 DU145 or PC3 cells were seeded in a T75 flask in 10 mL of complete medium and transfected the day after with 3 µg of the produced pGL4.13 *luc2*/SV40 (pGL4.13-A, pGL4.13-B, pGL4.13-C, pGL4.13-D). When performing the luciferase assay, cells were co-transfected also with 3 µg of the pRL-TK vector and with miRNA mimics at a final concentration of 40 nM. Cells were harvested at 48 hours after the transfection to perform different following assays. Depending on the experiments, the proportion of seeded cells and reagents was adapted according to the type of culturing support (T75 flasks, 60 mm dish, 6-well plates, or 24-well plates).
miRNA mimics and LNA/DNA mixmer ASOs were transiently transfected following an optimized protocol for the transfection of small molecules. 1 x 10^6 DU145 or PC3 cells were seeded in a T75 flask in 10 mL of complete medium and transfected the day after with miRNA mimics (*miR-140-5p*, *miR-132-3p*, *miR-518d-5p*, *miR-22-3p*, *miR-153-3p*) or with LNA/DNA mixmer ASOs (A, B, C, and D) and the relative controls (*miR-NEG* and ASO-NEG). Mix A was prepared using Opti-MEM™ medium (Thermo Fisher Scientific Inc., Waltham, MA, USA) and Lipofectamine™ 2000 (Thermo Fisher Scientific Inc., Waltham, MA, USA) with proportion of 150:1. Mix B was prepared using Opti-MEM™ medium and miRNA mimics or LNA/DNA mixmer ASOs at final concentration of 40 nM or 20 nM, respectively. When LNA/DNA mixmer ASOs were used as a combination, each was used at a final concentration of 20 nM; therefore, the control LNA/DNA mixmer ASO was used at a final concentration of 80 nM. Cells were harvested at 48 or 72 hours after the transfection and cell pellets were used for RNA and protein extraction. Sequences of miRNA mimics and LNA/DNA mixmer ASOs are reported in Supplementary Table S2.

Transcript decay. 5 x 10^5 DU145 cells were seeded in 60 mm dish in 5 mL complete medium per time point. The day after, cells were transfected with pGL4.13 *luc2*/SV40 vectors encoding the different chimeric

luciferases with the *NFYA* 3'UTRs (pGL4.13-A, pGL4.13-B, pGL4.13-C, pGL4.13-D). At 48 hours after transfection, cells were treated with the transcription inhibitor Actinomycin-D (Sigma-Aldrich, Saint Louis, MI, USA) at a final concentration of 10 µg/mL for 0, 2, 4, 8, and 24 hours. RNA of treated cells at different time points was extracted, reverse transcribed into cDNA and analyzed through qRT-PCR to evaluate the transcript decay of the chimeric luciferase. The half-life ($t_{1/2}$) was calculated as $\frac{ln(2)}{k}$ assuming first-order kinetics according to the exponential decay model $mRNA_t = mRNA_0 * e^{-kt}$. Data was normalized against the 0-hour time point for each tested gene and *c-MYC* was used as a positive control, since it is a known short-lived transcript [32].

Nucleus-cytoplasm fractionation. 1 x 10^6 DU145 cells were seeded in T75 flask in 10 mL complete medium per sample. The day after, cells were transfected with pGL4.13 *luc2*/SV40 vectors encoding the different chimeric luciferases with the *NFYA* 3'UTRs (pGL4.13-A, pGL4.13-B, pGL4.13-C, pGL4.13-D). At 48 hours after transfection, cells were harvested and resuspended in 300 µL of lysis buffer A (Tris-HCl pH 7.0 10 mM; NaCl 140 mM; MgCl2 1.5 mM; NP-40 0.5%; and deionized water), left on ice for 5 minutes, and then centrifuged. The supernatant (i.e., cytoplasm) was used for RNA extraction, while the pellet (i.e., nucleus) was washed three times with 300 µL of lysis buffer A and then resuspended in 300 µL of lysis buffer B (Tris-HCl pH 7.0 10 mM; NaCl 140 mM; MgCl2 1.5 mM; NP-40 0.5%; Tween-20 1%; deoxycholic acid 0.5%; and deionized water). After centrifugation, the pellet was used for RNA extraction and reverse transcription, and the cDNA of the nuclear and cytoplasmic fractions was analyzed through qRT-PCR to evaluate the percentage of chimeric luciferase expression in the two compartments. Cytoplasmic protein coding *GAPDH* and nuclear lncRNA *MALAT1* were used as positive controls for the correct fractionation. All reactions were carried out on ice and centrifugations at 1,000 g for 3 minutes at 4°C.

Analysis of subcellular RNA localization from the Villanueva *et al.* study. Bulk RNA-seq data from the "Localisation of RNA" (LoRNA) study of an untreated U2OS human cell line [33], was downloaded from the ENA repository (accession PRJEB49479). Fractions were annotated as "ER+mitochondria", "membraneless

cytosol", "nucleolus", and "unclear" based on the expression of the literature-curated markers of respective subcellular compartments, as described in the original study [33]. Of note, the "Nucleolus"-labelled fraction (#4) was enriched in both nucleolar and nuclear markers. We labeled it "Nucleolus" for consistency with the original study, but we interpreted it also as "nuclear fraction" in the main text. Short paired-end reads in .fastq format from 27 samples (3 bioreplicates per subcellular fraction, 8 fractions and Total Lysate) were uniformly processed with a custom developed snakemake [34] pipeline using snakemake v.8.25.3. Broken lines in .fastq files were sanitized and paired using the seqkit sana and seqkit pair tools [35]. Reads were further mapped with the STAR aligner [22] v.2.7.11b to the hg38 primary human genome, downloaded from the GENCODE website [36], without provided genome annotation.

STAR mapping was done with permissive parameters, to enable comprehensive detection of multi-mapped reads: STAR --runMode alignReads --alignEndsType Local --readFilesIn {input.fastq_1} {input.fastq_2} --readFilesCommand zcat --genomeLoad NoSharedMemory --outSAMtype BAM SortedByCoordinate --outFilterType BySJout --twopassMode None --outFilterMultimapNmax 500000000 --outSAMattrIHstart 0 --outSAMattributes All --limitOutSJcollapsed 5000000. Uniquely mapped reads were further retained to quantify gene expression in strand-specific mode with the featureCounts utility of subread package v.2.1.1 [37] and the following command-line parameters: featureCounts -p --countReadPairs -O --fraction -Q 255 -s 2 -B -C -P -d 1 -D 10000000000000 -t exon -g gene_id. As input, a custom .gtf annotation file was prepared based on the basic human GENCODE annotation v42 with python scripts, implemented within zavolan_pyutils package v1.0.1 [38]. Briefly, to avoid confounding effects from alternative isoform usage and non-coding RNA expression, this .gtf was constructed to retain exons with strong evidence of being constitutively included in transcripts. That is, exons that were not annotated as first or last in a transcript of that gene and were present in more than 80% of annotated transcript isoforms were used to construct a single artificial transcript isoform per gene. The resulting .gtf file included 124,588 exonic segments from 16,382 protein-coding genes, including the *NFYA* gene.

Obtained raw counts were further normalized across samples using the Bayesian inference procedure Sanity [39]. While established differential expression tools like DESeq2 [40]] and EdgeR [41, 42] fit generalized linear

models relying on a predefined condition design matrix, Sanity is a fully unsupervised Bayesian model. It infers the Maximum A Posteriori (MAP) estimate of the true latent log-transcription quotient (LTQ, i.e. the relative proportion of gene's RNA concentration in a total captured RNA content) and its associated posterior measurement variance for every individual gene and sample. This allows for flexible, post-hoc aggregation of data across conditions, enabling direct visualization of confidence intervals and statistical comparisons of mean condition expression levels without requiring multiple pairwise tests.

Sanity was originally designed for single-cell RNA-seq data, typically comprising thousands of cell-level transcriptome measurements. To adapt Sanity for bulk RNA-seq data with limited biological replicates and to introduce new functionality, we re-implemented the procedure in Python, termed "pySanity", provided within the zavolab_pyutils package. We validated that pySanity produces identical estimates to the original C++ implementation on simulated gene count data (see Supplementary Notebook "test_module" in the Zenodo repository). To robustly handle low-expressed genes in bulk data, we integrated an Empirical Bayes Shrinkage module into pySanity. Similar to DESeq2, this module derives a variance prior by fitting a LOWESS regression trend to the mean-variance relationship across all genes, and also applying a Chi-squared small-sample bias correction to accurately estimate overdispersion from limited replicates.

For principal component analysis (PCA), log2-transformed pySanity-normalized gene counts were utilized, with the lowest 30% of genes based on sample-mean expression being discarded.

Lastly, pySanity was extended to enable rigorous post-hoc analysis of alternative polyadenylation isoform usage. To strictly maintain the conditional independence of Poisson measurement errors assumed by the Bayesian model, relative usage of particular isoform was calculated relative to the aggregated counts of the "rest of the gene", allowing for exact propagation of Bayesian uncertainty when comparing isoform ratios across conditions. The input raw counts for each PAS isoform in samples were obtained with PAQR2 workflow [43], specifically its most recently published version [44]. The complete code of the described analysis is present in a dedicated branch "nfya" of the "APA_localization" github repository [45], release v0.1.0 [46].

Polysome fractionation. Actively growing DU145 cells were kept for 10 minutes at 37°C with culture medium added with 100µg/mL Cycloheximide and washed with ice-cold PBS with 10µg/mL Cycloheximide. The cells were scarpered in additional ice-cold PBS with 10µg/mL Cycloheximide and centrifuged at 400 g for 3 minutes at 4°C. Cell pellets were lysed in lysis buffer (50mM Tris HCl pH 7.5, 100mM NaCl, 30mM $MgCl_2$, 0.1% NP-40, 100µg/mL Cycloheximide, 40U/mL RNase inhibitor, and protease inhibitor 1X in nuclease free water) and let stand on ice for 30 minutes. Cytoplasmic lysates were clarified at 14,000 g for 10 min at 4°C. Equal amounts of RNA were loaded on 10-50% sucrose gradient dissolved in 50mM Tris HCl pH 7.5, 100mM NaCl, 12mM $MgCl_2$, and 1mM DTT, and centrifuged at 39,000 rpm for 3 h and 30 minutes in a SW41 Beckman rotor at 4°C. The gradient was then analyzed by continuous flow absorbance at 254 nm, recorded by UV-BioLogic LP device (Bio-Rad) at 1mL/min. Fractions were collected and pooled together to get the soluble (before the 40S peak), sub-polysomes (from the 40S to the 80S peak), light polysomes, and heavy polysomes pooled fractions. Samples were incubated with proteinase K (20 µg/µL per fraction) and SDS 1% for 1 h and 30 minutes at 37°C, and RNA was extracted by phenol/chloroform/isoamyl alcohol method.

Cell proliferation assay. 6 x 10^4 WT and ΔPAS cells or ASO-treated cells were plated in a 6-well plate in 3 mL complete medium, in triplicates. Whereas 1.5 x 10^5 DU145 cells were seeded in a 6-well plate in 3 mL complete medium, in triplicates, and the day after were transfected with LNA/DNA mixmer ASOs. Viable cell number was counted at 24, 48, 72, and 96 hours after seeding/transfection with the TC20™ Automated Cell Counter (Bio-Rad, Hercules, CA, USA). Data were normalized against WT or ASO-NEG at 48 or 24 hours, respectively.

Colony formation assay. 500 WT and ΔPAS cells were plated in 6-well plates in 3mL complete medium, in triplicate. After 14 days, plates were washed twice with PBS and fixed/stained with a Crystal Violet solution (Crystal violet 0.5%, Formaldehyde 1%, Methanol 1%, in PBS) for 30 minutes at room temperature. Colonies were counted using ImageJ software.

Wound-healing assay. WT and ΔPAS cells or ASO-treated cells were cultured until they reached 90–95% confluency. Cells were washed with PBS and scratched to create the wound line. The rate of wound-healing was monitored at 0h, 4h, 8h, 24h, and 32h after scratch by taking images with a microscope (HCS ImageXpress Micro Confocal) and the average gap width was calculated using ImageJ software.

In vivo experiments

***Animal care and handling*.** The Tg(kdrl:DsRed) transgenic line was maintained and handled in compliance with national guidelines on animal welfare (Italian Decree 4th March 2014, n°26). Adult fish were used exclusively for breeding; xenotransplantation and imaging experiments were performed on zebrafish larvae.

***Microinjection of zebrafish larvae with human prostate cancer cells*.** WT1 and ΔPAS1 cells were transduced with Tet-Off-H2B-GFP lentiviral supernatants generated by transfecting HEK293T cells (ATCC, Cat# CRL-3216, Manassas, VA, USA) using the calcium phosphate transfection method. Following transduction, GFP expression was assessed after 24 hours and GFP high cells were selected by FACS-sorting and subsequently expanded in vitro to obtain stable fluorescently-labeled GFP+ WT1 and ΔPAS1 cells, which were then maintained under standard culture conditions at 37°C in a humidified incubator with 21% O_2 and 5% CO_2. Cells were cultured in RPMI Medium (EuroClone), supplemented with 10% South American fetal bovine serum (FBS, EuroClone), 2 mM L-glutamine (EuroClone), 25 mM HEPES and 100 U/mL penicillin-streptomycin (Life Technologies). Cells were passaged twice a week upon reaching 80-90% confluence.

On the day of transplantation, PCa cells were harvested, resuspended in complete medium at a concentration of 106 cells/ml and kept on ice throughout the experiment. Xenotransplantation was performed on 2-day-old Tg(kdrl:DsRed) transgenic larvae [47] maintained in standard E3 embryo medium supplemented with 0.003% (w/v) phenylthiourea (PTU, Sigma-Aldrich) to inhibit melanogenesis.

Microinjection setup. Customized microinjection needles were prepared by pulling GC100T-15 borosilicate glass capillaries (1.0 mm outer diameter / 0.78 mm inner diameter, Harvard Apparatus) using a P-97

micropipette puller (Sutter Instruments) with the following parameters: heat: 505; pull: 60; velocity: 60; time: 100; air pressure: 300. Upon loading with the cell suspension, needles were inserted into a mechanical micromanipulator (MN-153, Narishige Group) connected to an electric microinjector (IM-401, Narishige Group).

Microinjection procedure. Zebrafish larvae were anesthetized in 0.016% tricaine methanesulfonate (MS-222, Sigma-Aldrich) dissolved in E3 medium with PTU and positioned laterally on a 3% agarose bed under an Olympus SZX9 stereomicroscope. The injection droplet volume was determined by adjusting injection time and pressure, using a micrometric scale positioned in the stereomicroscope eyepiece. Each larva was injected into perivitelline space with 5 nL of cell suspension, corresponding to ~500 cells. Injected larvae were allowed to recover for two hours at 28°C and then screened under a Nikon SMZ25 fluorescence stereomicroscope to exclude dead or mis-injected individuals (e.g., injected into the yolk sac or with an incorrect volume of cells). Correctly injected larvae were then kept at 34°C.

Imaging procedures

Stereomicroscopy imaging. Stereomicroscopy imaging on live larvae was performed to assess tumor engraftment and quantify metastases at 1 and 3 days post injection (dpi); live xenografts were arrayed in 48-well plates, immobilized with a few drops of 0.016% tricaine in E3 water + PTU and observed using a Nikon SMZ25 stereomicroscope equipped with Nis Elements AR software (v. 5.21.02) and a 1x/N.A. 0.15-W.D. 60 mm Planapo objective. Fluorescence excitation was provided by a Lumencor SolA SE U-nIR illuminator, using GFP and RFP filters to detect GFP and DsRed signals from PCa cells and zebrafish endothelial cells, respectively. Images were acquired with a sCMOS Orca Hamamatsu FL4.0 LT camera at 40x zoom (1 dpi) or 30x (3 dpi) zoom.

Confocal microscopy imaging. Confocal imaging was performed on 3 dpi larvae euthanized in 0.4% tricaine and fixed for two hours in 4% paraformaldehyde (Sigma-Aldrich). Fixed larvae were washed in PBS and mounted laterally on glass slides in glycerol mounting medium. Imaging was performed on a Leica SP8 FSU AOBS confocal microscope using resonant scanning; whole-larva z-stacks and tile scans were acquired with a

10x/0.3 dry objective. Samples were simultaneously excited with 488 nm and 561 nm lasers to acquire signals from GFP+ PCa cells and DsRed+ zebrafish endothelial cells; transmitted light images were also collected. The stitching was performed using the Mosaic Merge function in Leica LasX software (v. 3.5.5.19976) and maximum intensity projections were generated with the corresponding LasX tools.

For clarity, all images underwent brightness/contrast adjustments and noise filtering in Fiji (version 2.16.0/1.54p).

Analyses of engraftment and metastatic potential of prostate cancer cells.

Xenotransplantation experiments were performed using different PCa cell preparations and multiple batches of larvae; both the conditions (WT1 and ΔPAS1 cells) were tested within the same experimental round. To assess the engraftment and metastatic potential of PCa cells, we quantified the following parameters.

Engraftment rate. The engraftment rate—defined as the proportion of animals exhibiting local growth following cell injection—was calculated for 3 dpi larvae by visual inspection under a Nikon SMZ25 stereomicroscope. Larvae were scored as engrafted when a primary tumor (PT) was detectable at the injection site. Statistical difference between groups was assessed using two-sided Fisher exact test.

Incidence of metastasis. The incidence of metastasis—defined as the ratio of engrafted larvae bearing at least one metastasis at 1 or 3 dpi—was determined by scoring the number of engrafted larvae showing at least one cluster of ≥5 extravasated cells. Pairwise comparisons between groups were performed using two-sided Fisher exact tests.

Average number of metastases. The average number of metastases—defined as the ratio of the total number of metastases observed to the total number of engrafted larvae—was calculated by visual inspection of 1 or 3 dpi engrafted larvae under a Nikon SMZ25 stereomicroscope and manually counting the number of clusters of ≥ 5 extravasated cells. Statistical difference between groups was determined pairwise by two-sided Mann-Whitney tests.

Statistical analyses were performed in GraphPad Prism (version 10.4.2) using data pooled from at least four independent experiments. A p-value < 0.05 was considered statistically significant.

REFERENCES

1. Fazal FM, Han S, Parker KR, Kaewsapsak P, Xu J, Boettiger AN, et al. Atlas of Subcellular RNA Localization Revealed by APEX-seq. Cell. 2019;178:473-490.e26. doi:10.1016/j.cell.2019.05.027.

2. Robinson JT, Thorvaldsdottir H, Turner D, Mesirov JP. igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). Bioinformatics. 2022;39:btac830. doi:10.1093/bioinformatics/btac830.

3. Hong W, Ruan H, Zhang Z, Ye Y, Liu Y, Li S, et al. APAatlas: decoding alternative polyadenylation across human tissues. Nucleic Acids Res. 2020;48:D34–9. doi:10.1093/nar/gkz876.

4. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45:580–5. doi:10.1038/ng.2653.

5. Feng X, Li L, Wagner EJ, Li W. TC3A: The Cancer 3’ UTR Atlas. Nucleic Acids Res. 2018;46 Database issue:D1027–30. doi:10.1093/nar/gkx892.

6. Weinstein JN, Collisson EA, Mills GB, Shaw KM, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer Analysis Project. Nat Genet. 2013;45:1113–20. doi:10.1038/ng.2764.

7. Xia Z, Donehower LA, Cooper TA, Neilson JR, Wheeler DA, Wagner EJ, et al. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3’-UTR landscape across seven tumour types. Nat Commun. 2014;5:5274. doi:10.1038/ncomms6274.

8. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modeling of anticancer drug sensitivity. Nature. 2012;483:603–7. doi:10.1038/nature11003.

9. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a Cancer Dependency Map. Cell. 2017;170:564-576.e16. doi:10.1016/j.cell.2017.06.010.

10. Londero M, Gallo A, Cattaneo C, Ghilardi A, Ronzio M, Del Giacco L, et al. NF-YAl drives EMT in Claudinlow tumours. Cell Death Dis. 2023;14:65. doi:10.1038/s41419-023-05591-9.

11. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323. doi:10.1186/1471-2105-12-323.

12. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet. 2007;39:1181–6. doi:10.1038/ng1007-1181.

13. Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, et al. NCBI’s Database of Genotypes and Phenotypes: dbGaP. Nucleic Acids Res. 2014;42:D975–9. doi:10.1093/nar/gkt1211.

14. Heath AP, Ferretti V, Agrawal S, An M, Angelakos JC, Arya R, et al. The NCI Genomic Data Commons. Nat Genet. 2021;53:257–62. doi:10.1038/s41588-021-00791-5.

15. Dataset from The Cancer Genome Atlas project. Accession id: phs000178.v11.p8. https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000178.v11.p8 Accessed 22 September 2025.

16. National Cancer Institute (NCI). GDC Data Transfer Tool. https://github.com/NCI-GDC/gdc-client. Accessed 22 September 2025.

17. recount3: uniformly processed RNA-seq. https://rna.recount.bio/. Accessed 10 March 2026.

18. Wilks C, Zheng SC, Chen FY, Charles R, Solomon B, Ling JP, et al. recount3: summaries and queries for large-scale RNA-seq expression and splicing. Genome Biol. 2021;22:323. doi:10.1186/s13059-021-02533-6.

19. Wilks C, Ahmed O, Baker DN, Zhang D, Collado-Torres L, Langmead B. Megadepth: efficient coverage quantification for BigWigs and BAMs. Bioinformatics. 2021;37:3014–6. doi:10.1093/bioinformatics/btab152.

20. Dataset from the Stand Up To Cancer East Coast Prostate Cancer Research Group. Accession id: phs000915.v2.p2. https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000915.v2.p2. Accessed 16 January 2026

21. National Center for Biotechnology Information (NCBI). SRA Toolkit. https://github.com/ncbi/sra-tools. Accessed 3 June 2025.

22. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21. doi:10.1093/bioinformatics/bts635.

23. Ye C, Long Y, Ji G, Li QQ, Wu X. APAtrap: identification and quantification of alternative polyadenylation sites from RNA-seq data. Bioinformatics. 2018;34:1841–9. doi:10.1093/bioinformatics/bty029.

24. Nassar LR, Barber GP, Benet-Pagès A, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2023 update. Nucleic Acids Res. 2022;51:D1188–95. doi:10.1093/nar/gkac1072.

25. Herrmann CJ, Schmidt R, Kanitz A, Artimo P, Gruber AJ, Zavolan M. PolyASite 2.0: a consolidated atlas of polyadenylation sites from 3' end sequencing. Nucleic Acids Res. 2020;48:D174–9. doi:10.1093/nar/gkz918.

26. Wang R, Nambiar R, Zheng D, Tian B. PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. Nucleic Acids Res. 2018;46:D315–9. doi:10.1093/nar/gkx1000.

27. Moon Y, Herrmann CJ, Mironov A, Zavolan M. PolyASite v3.0: a multi-species atlas of polyadenylation sites inferred from single-cell RNA-sequencing data. Nucleic Acids Res. 2025;53:D197–204. doi:10.1093/nar/gkae1043.

28. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50. doi:10.1073/pnas.0506580102.

29. White EZ, Pennant NM, Carter JR, Hawsawi O, Odero-Marah V, Hinton CV. Serum deprivation initiates adaptation and survival to oxidative stress in prostate cancer cells. Sci Rep. 2020;10:12505. doi:10.1038/s41598-020-68668-x.

30. Mitra M, Ho LD, Coller HA. An in vitro model of cellular quiescence in primary human dermal fibroblasts. Methods Mol Biol. 2018;1686:27–47. doi:10.1007/978-1-4939-7371-2_2.

31. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. 2007;35 Web Server issue:W71–4. doi:10.1093/nar/gkm306.

32. Miller DM, Thomas SD, Islam A, Muench D, Sedoris K. c-Myc and Cancer Metabolism. Clin Cancer Res. 2012;18:5546–53. doi:10.1158/1078-0432.CCR-12-0977.

33. Villanueva E, Smith T, Pizzinga M, Elzek M, Queiroz RML, Harvey RF, et al. System-wide analysis of RNA and protein subcellular localization dynamics. Nat Methods. 2024;21:60–71. doi:10.1038/s41592-023-02101-9.

34. Köster J, Rahmann S. Snakemake--a scalable bioinformatics workflow engine. Bioinformatics. 2012;28:2520–2. doi:10.1093/bioinformatics/bts480.

35. Shen W, Sipos B, Zhao L. SeqKit2: A Swiss army knife for sequence and alignment processing. Imeta. 2024;3:e191. doi:10.1002/imt2.191.

36. Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 2019;47 Database issue:D766–73. doi:10.1093/nar/gky955.

37. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013;41:e108. doi:10.1093/nar/gkt214.

38. Mironov A. zavolab_pyutils-v1.0.0. 2026. doi:10.5281/zenodo.19483710.

39. Breda J, Zavolan M, van Nimwegen E. Bayesian inference of gene expression states from single-cell RNA-seq data. Nat Biotechnol. 2021;39:1008–16. doi:10.1038/s41587-021-00875-x.

40. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. doi:10.1186/s13059-014-0550-8.

41. Chen Y, Chen L, Lun ATL, Baldoni PL, Smyth GK. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. Nucleic Acids Res. 2025;53:gkaf018. doi:10.1093/nar/gkaf018.

42. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40. doi:10.1093/bioinformatics/btp616.

43. Gruber AJ, Schmidt R, Ghosh S, Martin G, Gruber AR, van Nimwegen E, et al. Discovery of physiological and cancer-related regulators of 3’ UTR processing with KAPAC. Genome Biol. 2018;19:44. doi:10.1186/s13059-018-1415-3.

44. Mironov A, Franchitti L, Ghosh S, Ritz M-F, Hutter G, De Bortoli M, et al. Leveraging multi-omics data to infer regulators of mRNA 3’ end processing in glioblastoma. Front Mol Biosci. 2024;11:1363933. doi:10.3389/fmolb.2024.1363933

45. Zavolan Lab. APA_localization (nfya branch). https://github.com/zavolanlab/APA_localization.

46. magmir71. zavolanlab/APA_localization: v0.1.0 - initial pre-release, with NFYA project materials. 2026. doi:10.5281/zenodo.19482095.

47. Jin S-W, Beis D, Mitchell T, Chen J-N, Stainier DYR. Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. Development. 2005;132:5199–209. doi:10.1242/dev.02087.