nature genetics

SUPPLEMENTARY INFORMATION

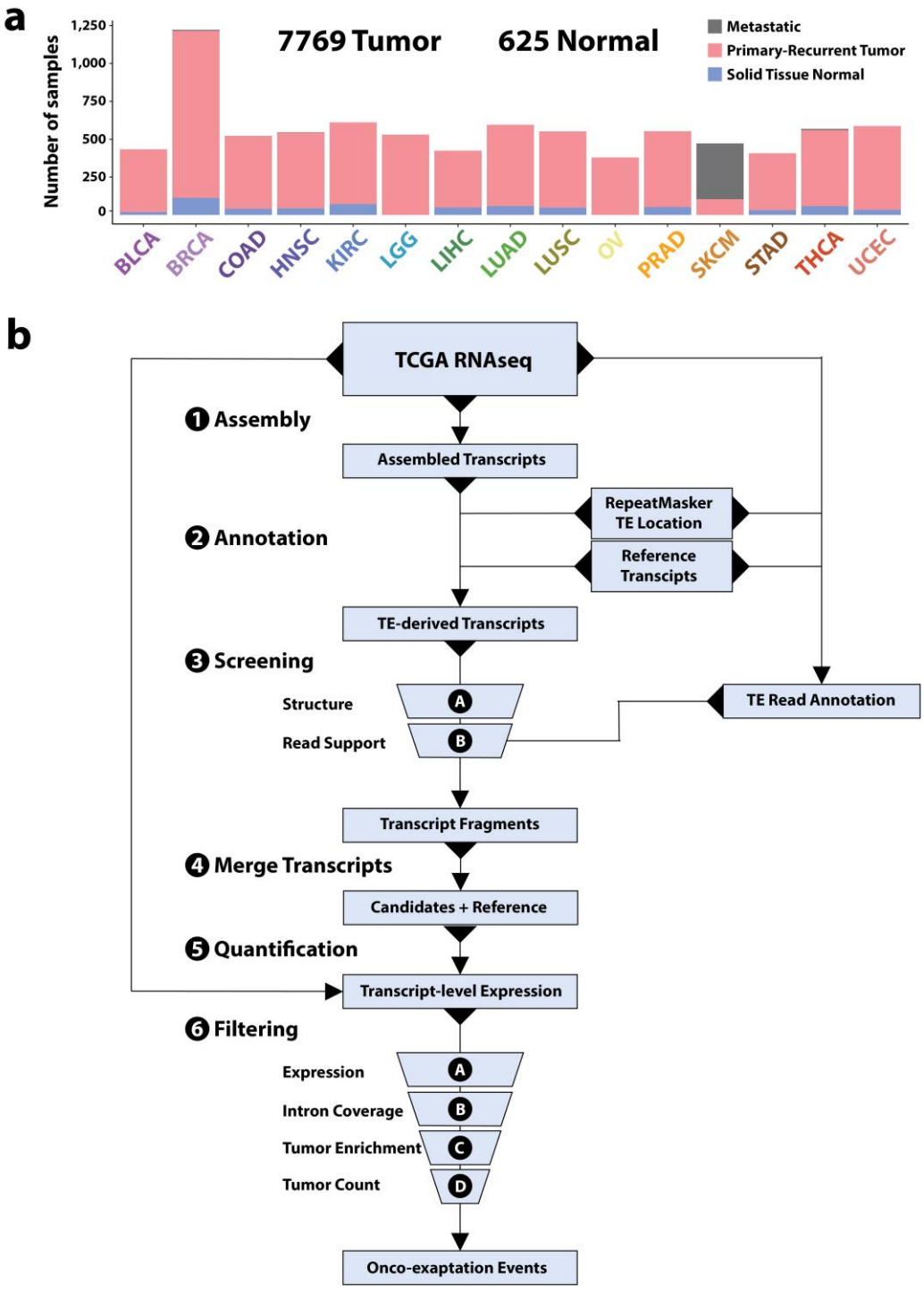
https://doi.org/10.1038/s41588-019-0373-3

In the format provided by the authors and unedited.

Transposable elements drive widespread expression of oncogenes in human cancers

Hyo Sik Jang[1,2,5], Nakul M. Shah[1,2,5], Alan Y. Du[1,2], Zea Z. Dailey[1,2], Erica C. Pehrsson[1,2], Paula M. Godoy[1,2], David Zhang[1,2], Daofeng Li[1,2], Xiaoyun Xing[1,2], Sungsu Kim[1,3], David O'Donnell[1,2,4], Jeffrey I. Gordon[1,2,4] and Ting Wang[1,2*]

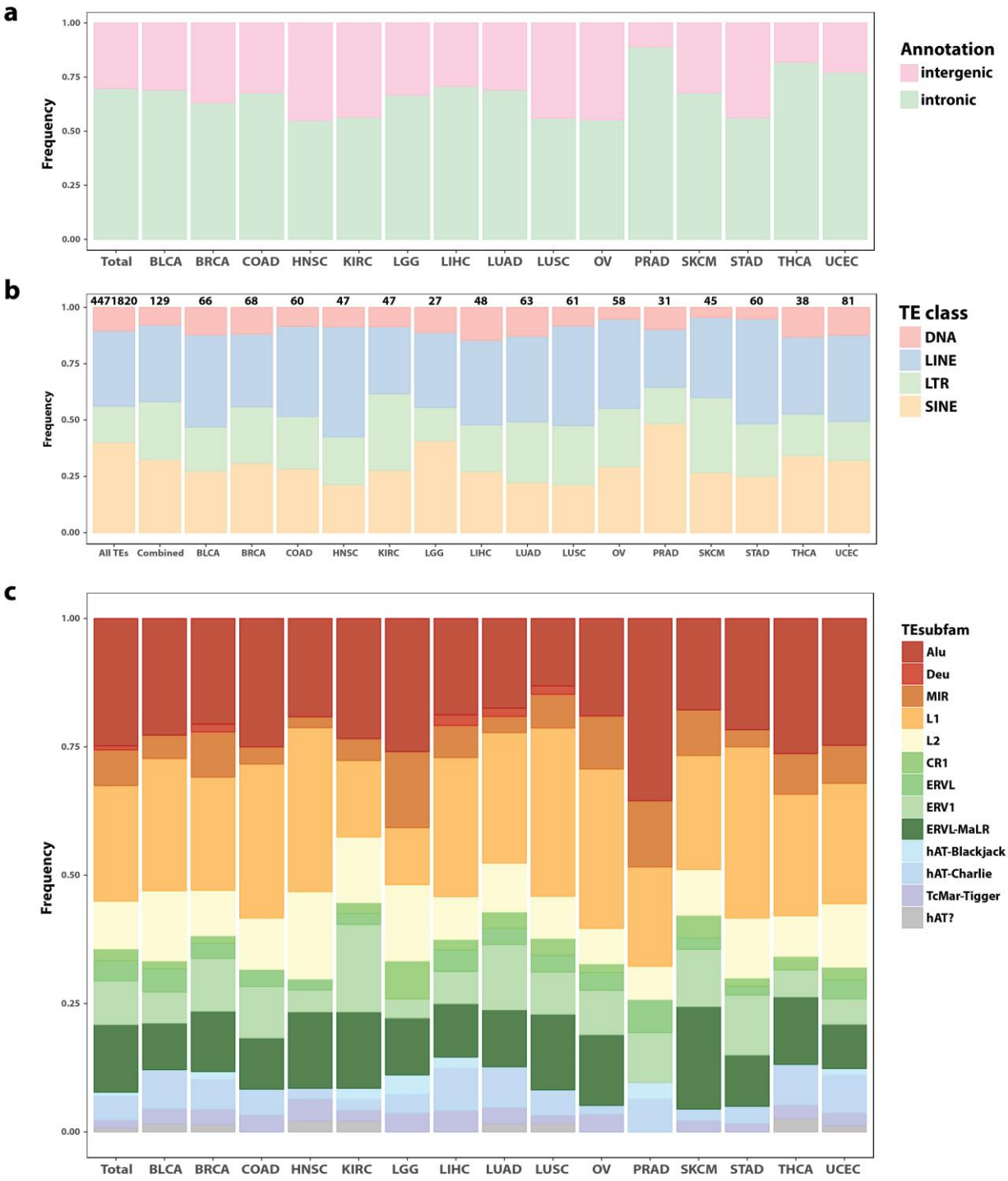
[1]Department of Genetics, Washington University School of Medicine, St Louis, MO, USA. [2]The Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St Louis, MO, USA. [3]Hope Center for Neurological Disease, Washington University School of Medicine, St Louis, MO, USA. [4]Center for Gut Microbiome and Nutrition Research, Washington University School of Medicine, St Louis, MO, USA. [5]These authors contributed equally: Hyo Sik Jang and Nakul M. Shah. *e-mail: twang@wustl.edu



Supplementary Figure 1

RNA-seq computational pipeline detects numerous TE onco-exaptation events in 15 cancer types from TCGA.

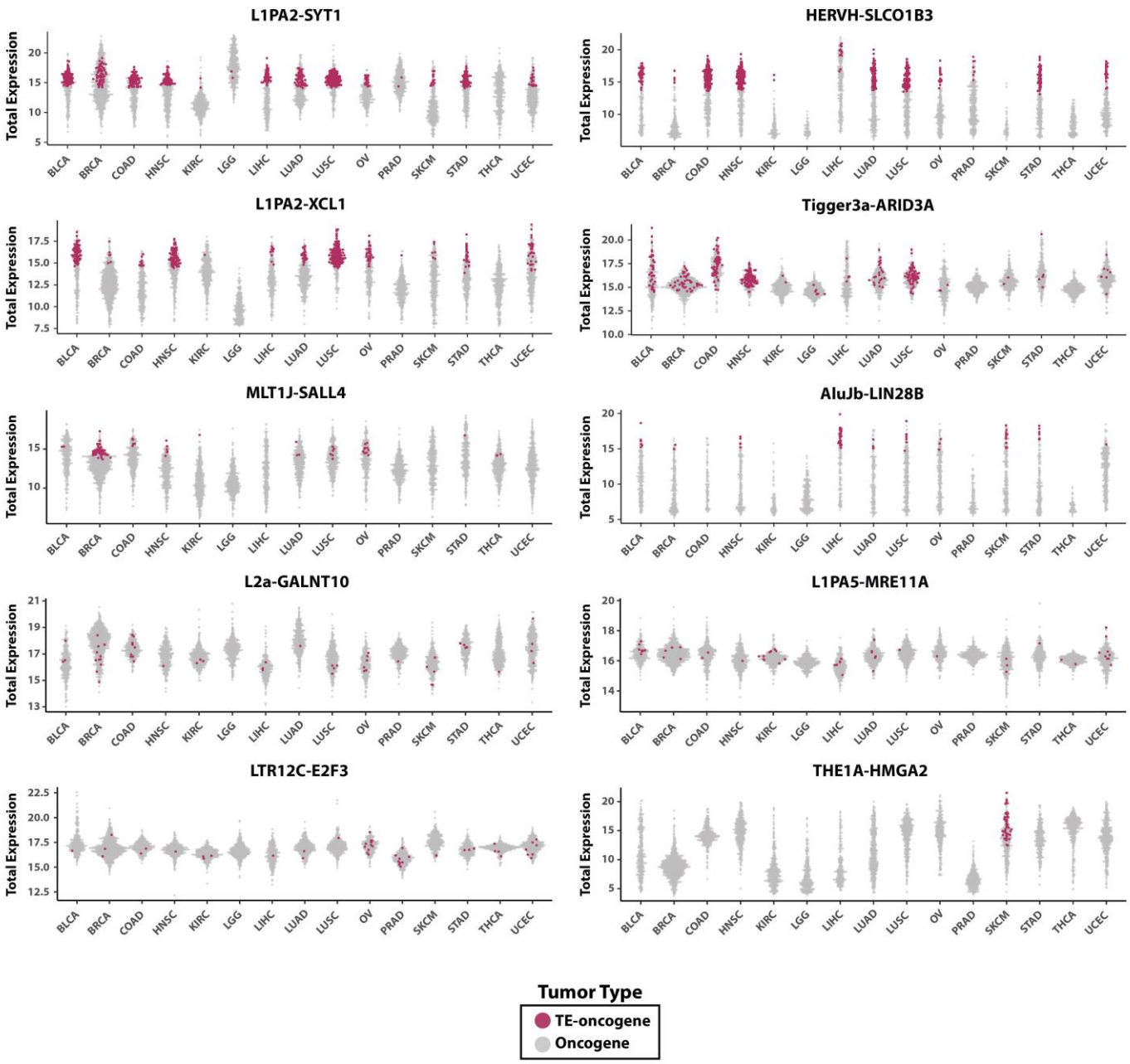
a, Number of cases from various cancer types that were analyzed**. b,** Schematic of the computational pipeline describing how RNA-seq from TCGA was processed to identify onco-exaptation candidates.



Supplementary Figure 2

TE locations and annotations that are implicated in onco-exaptation events.

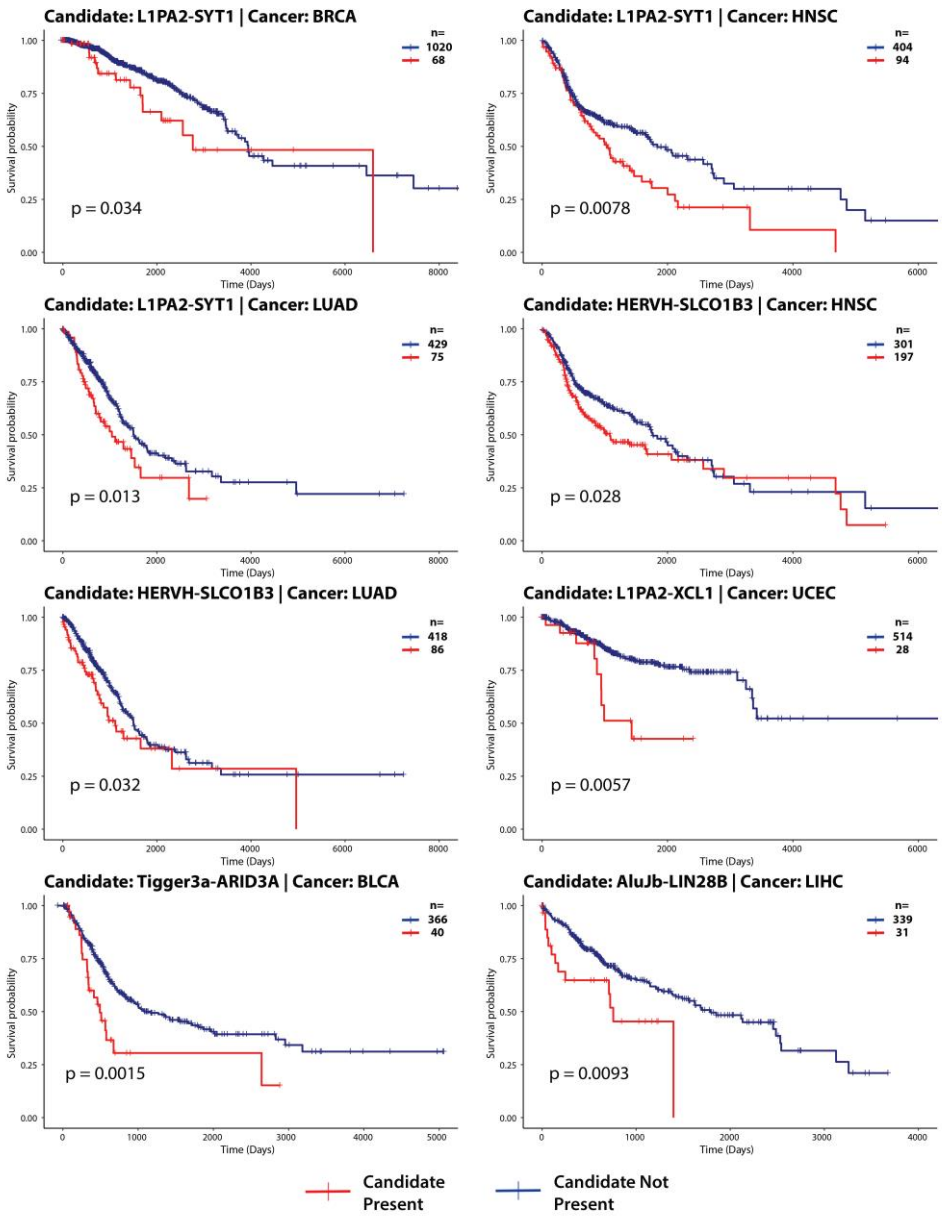
a, The genomic locations of TEs that act as cryptic promoters for oncogenes across different cancer types. **b,** Distribution of TE classes across cancer types. Total number of unique TEs that contribute to onco-exaptation events are labeled on top. **c,** Distribution of each TE family that contributed to onco-exaptation across 15 cancer types.



Supplementary Figure 3

Oncogene expression profiles of the top ten onco-exaptation candidates across cancer types.

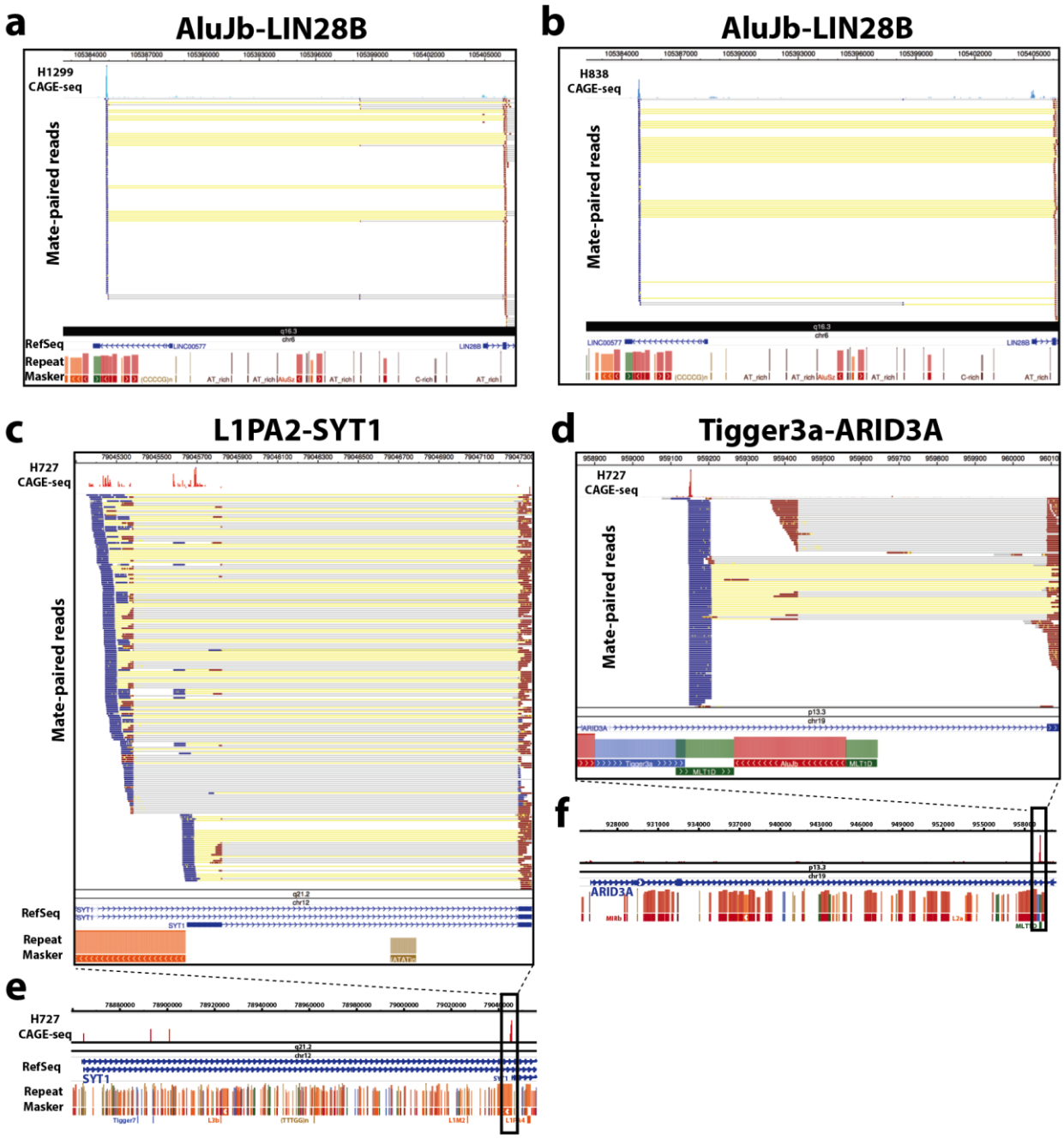
Oncogene expression of each tumor with and without an onco-exaptation event. Each grey dot represents a tumor while the red dots reveal whether the tumor is predicted to have an onco-exaptation event. Total expression is represented as $\log_2$(FPKM-UQ) provided by TCGA GDC (https://portal.gdc.cancer.gov/).



Supplementary Figure 4

Overall survival impact of onco-exaptation candidates.

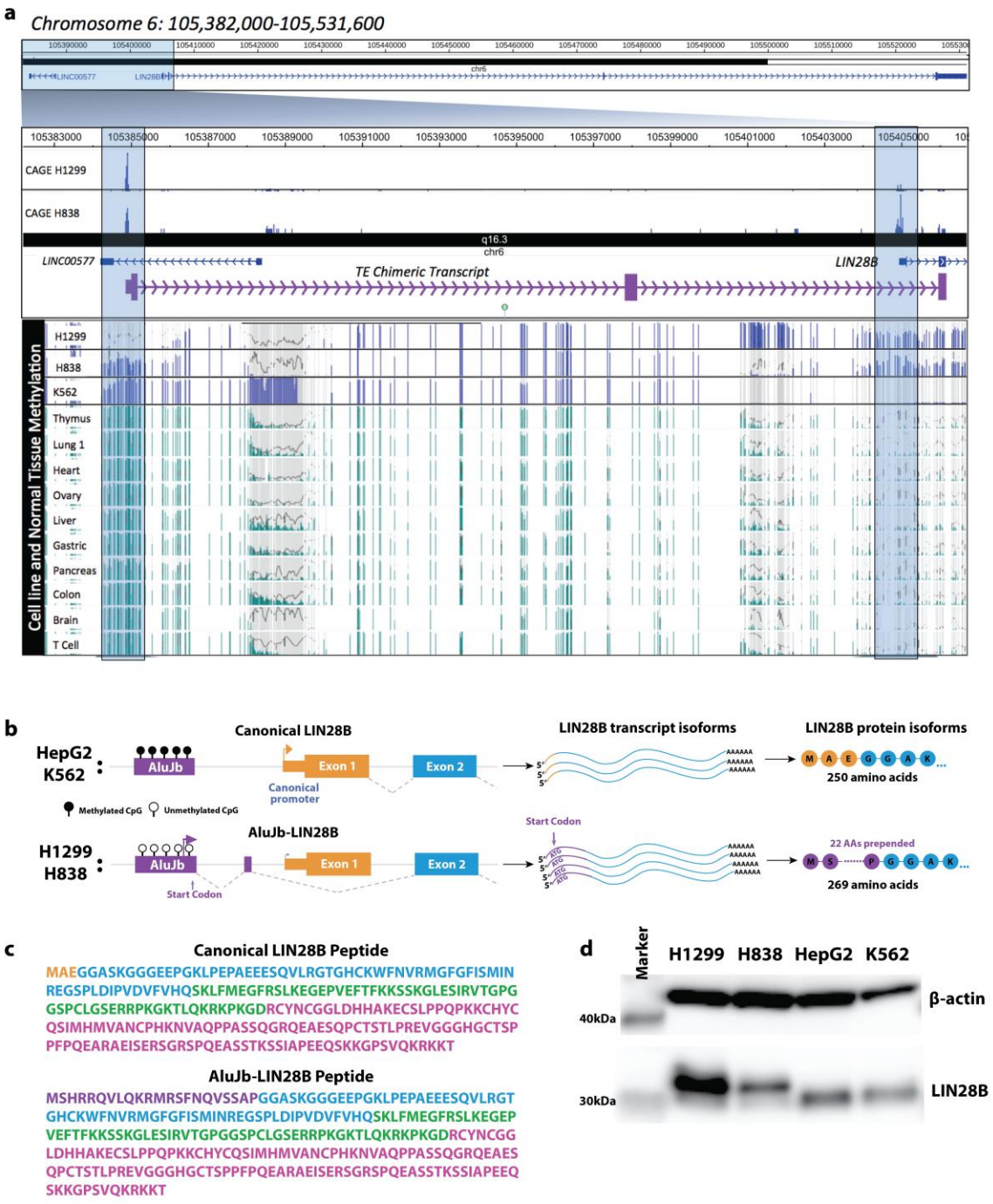
Kaplan-Meier Curves for the 8 examples of where a top 10 candidate was significantly prognostic in a cancer ($p<0.05$) based on log-rank statistical test (two-sided). The red line in each graph represents patients where the candidate was found to be present, and in blue line represents all the patients where the candidate was not detected. All were found to negatively impact overall survival. The number of biologically independent patients is listed within each plot.



Supplementary Figure 5

Transcription start site verification of onco-exaptation candidates in H727 lung cancer cell line.

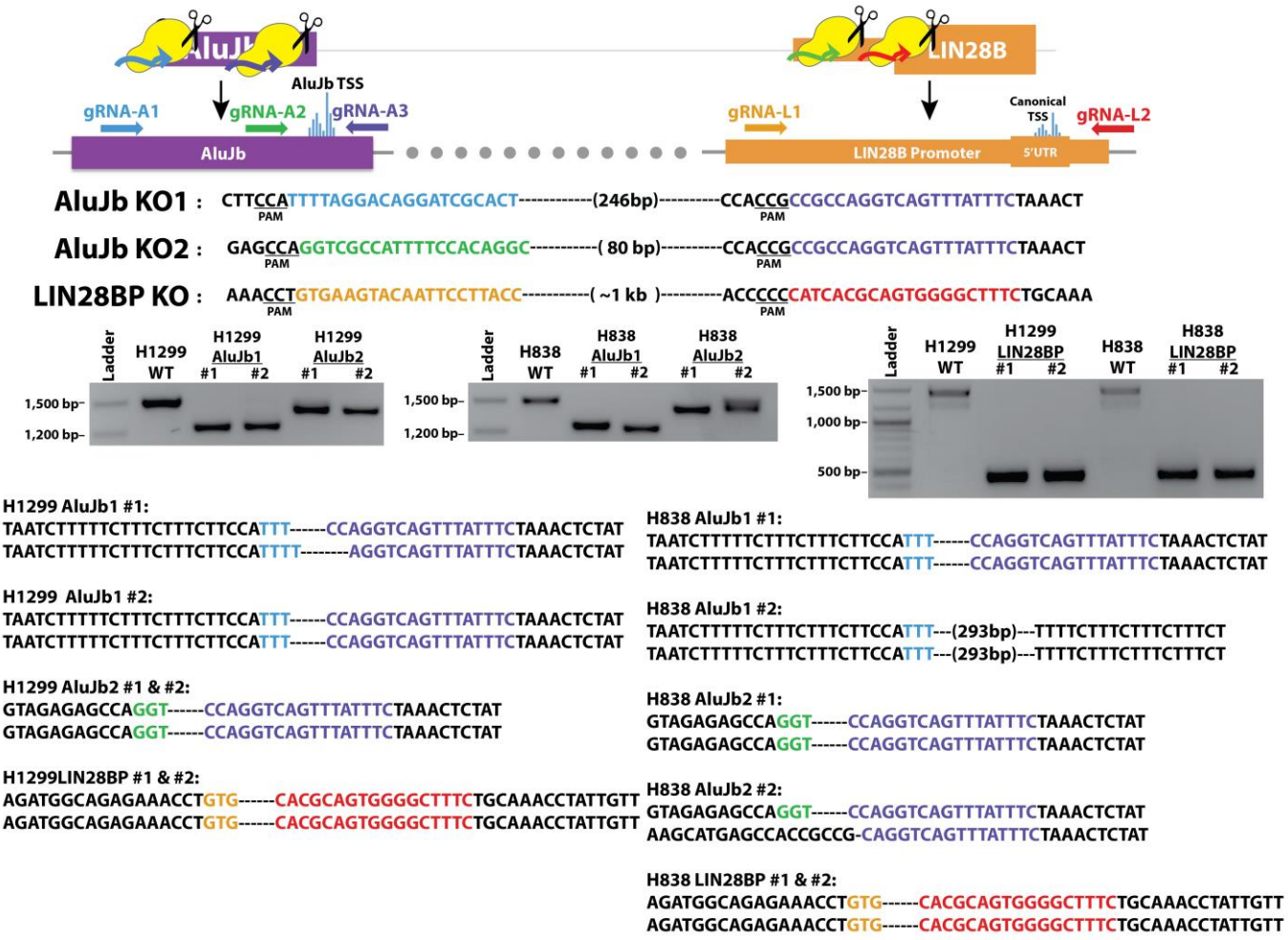
a, WashU Epigenome browser (http://epigenomegateway.wustl.edu/browser/) view of CAGE-seq and mate-paired reads where the forward read initiates from AluJb and the reverse read ends in the gene body of LIN28B in H1299 and **b,** H838. **c,** WashU Epigenome browser view of H727 CAGE-seq and mate-paired reads that where the forward read from L1PA2 and the reverse read ends in the gene body of SYT1. **d,** WashU Epigenome browser view of H727 CAGE-seq and mate-paired reads where the forward read initiates from Tigger3a/MLT1D and the reverse read ends in the gene body of ARID3A. **e,** WashU Epigenome browser view of H727 CAGE-seq over SYT1 promoters. **f,** WashU Epigenome browser view of H727 CAGE-seq over ARID3A promoter.



Supplementary Figure 6

The AluJb TE is methylated in somatic tissue and is epigenetically dysregulated in lung cancer cell lines.

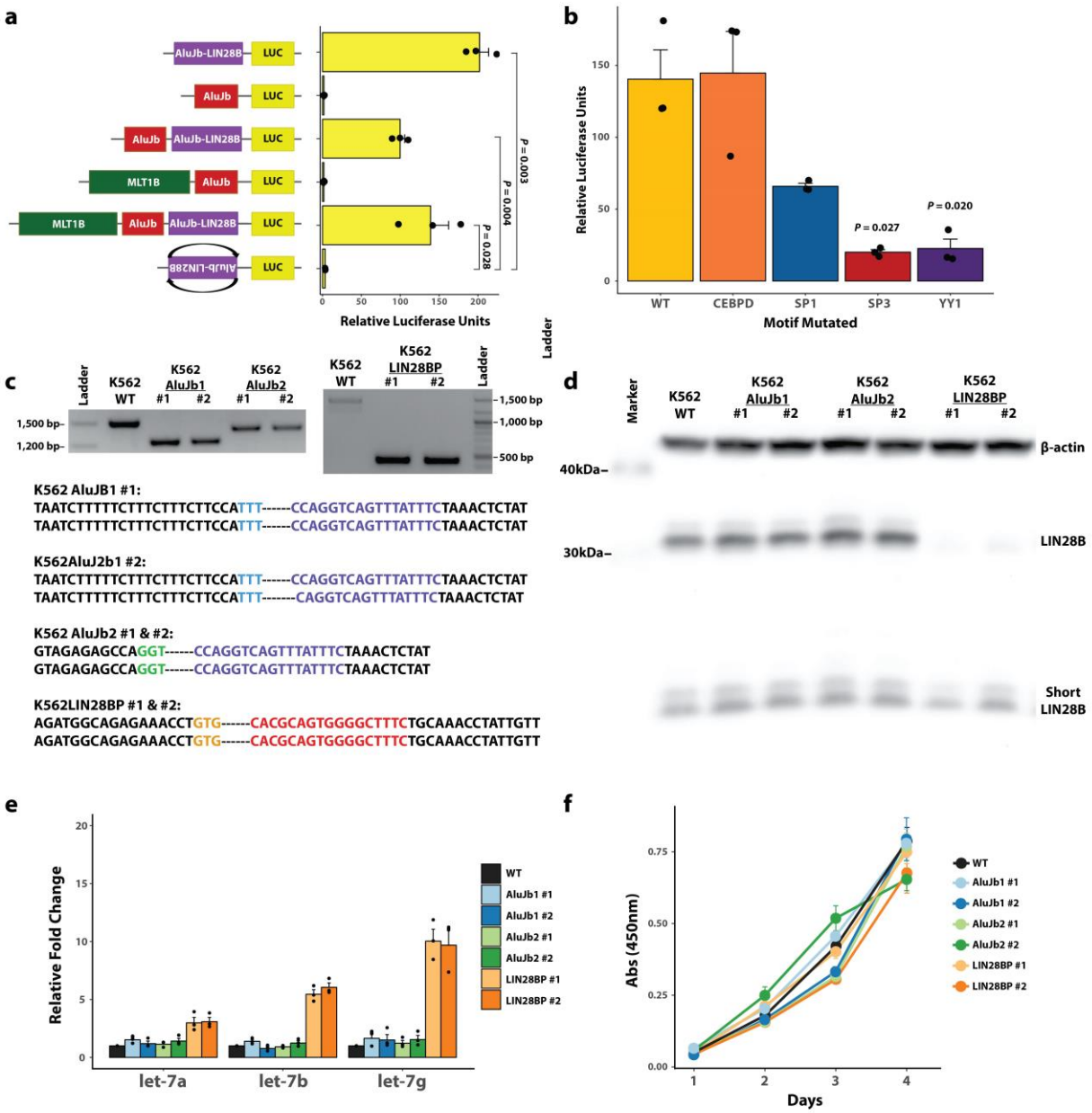
a, DNA methylome profiles of multiple somatic tissues from the Roadmap Epigenomics Project (http://www.roadmapepigenomics.org/) are displayed on the WashU Epigenome Browser (http://epigenomegateway.wustl.edu/browser/). **b,** Schematics showing DNA methylation levels of AluJb in different cancer cell lines. An alternative start codon present in AluJb generates a chimeric LIN28B peptide that lacks 3 amino acids contributed by exon 1, but has 22 novel amino acids prepended. **c,** Predicted amino acid sequence of AluJb-LIN28B protein. **d,** Cropped Western blot (repeated twice with similar results) representing the size difference between the AluJb-LIN28B protein and canonical LIN28B protein.



Supplementary Figure 7

Genotypes of AluJb-P and LIN28BP CRISPR-deleted clones.

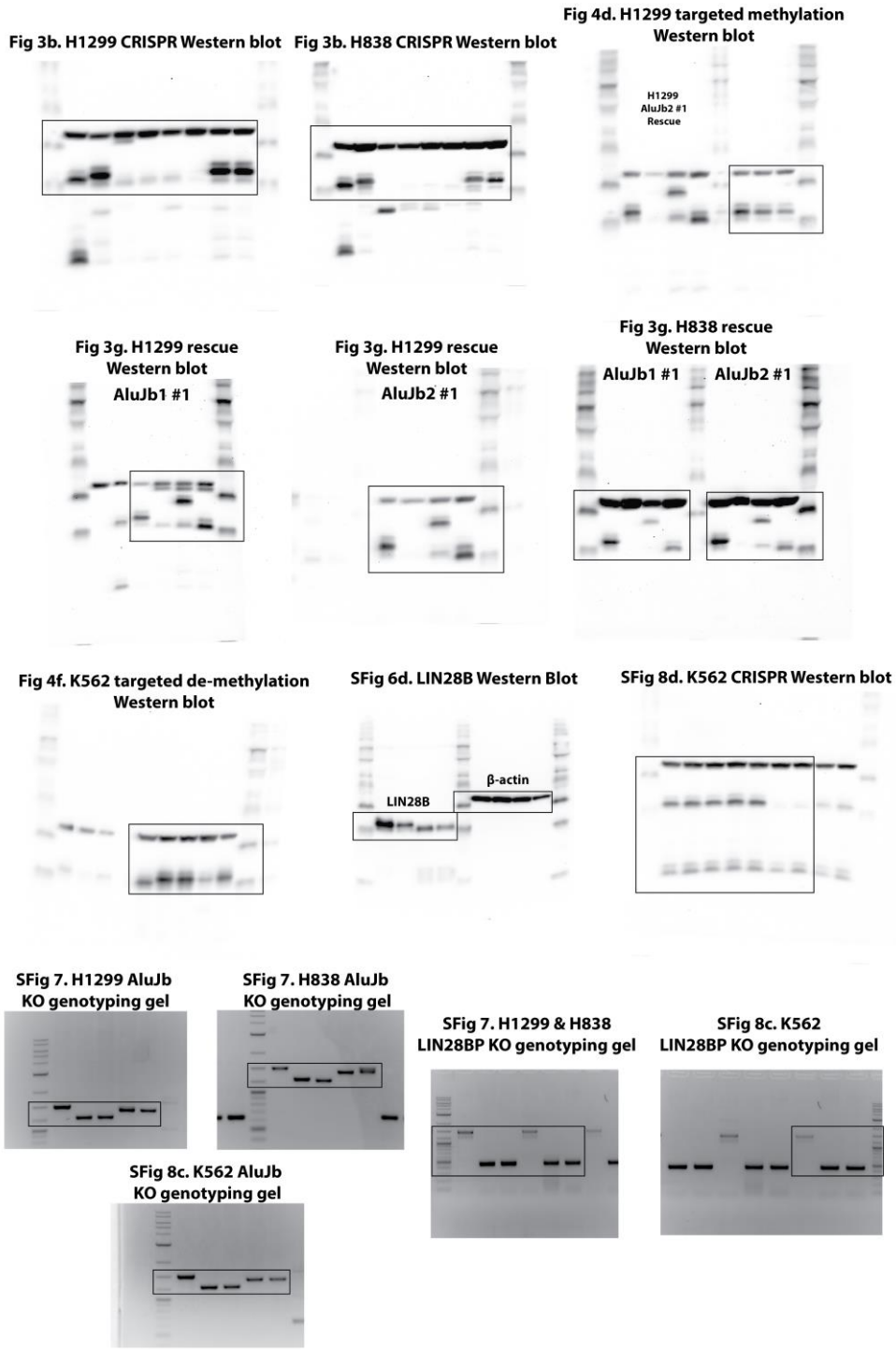
The CRISPR-mediated genetic breaks are illustrated for H1299 (left) and H838 (right) CRISPR clones. gRNA sequences are illustrated with various colors. AluJb1 KO clones were generated by deleted genomic region between gRNA-A1 and gRNA-A3. AluJb2 KO clones were generated by deleted genomic region between gRNA-A2 and gRNA-A3. LIN28BP KO clones were generated by deleted genomic region between gRNA-L1 and gRNA-L2. Two independent clones were profiled for each set of CRISPR KOs. Two independent clones were profiled for each set of CRISPR KOs. The genotyping gel results were replicated twice in independent experiments.



Supplementary Figure 8

K562 control experiments for AluJb-LIN28B candidate validation.

a, Promoter luciferase (n = 3 independent experiments) results illustrating transcriptional activity of various TE arrangements in K562. Welch's t-test was performed against reverse complement AluJb sequence (limited activity). **b,** Luciferase assays (n = 3 independent experiments) for mutagenized transcription factor motifs in K562. Relative luciferase activity for mutagenized vectors were compared to wild-type AluJb-P. **c,** Genotypes of K562 CRISPR KO clones for AluJb-P and LIN28BP deletions. Genotype check was repeated twice with similar results. **d,** Cropped Western blot for LIN28B in K562 CRISPR KO clones. K562 also expresses a smaller isoform of LIN28B that is not present in H1299 and H838. This experiment was repeated twice with similar results. **e,** Relative let-7a, let-7b and let-7g miRNA levels compared to wild-type in CRISPR-knockout clones of K562 as measured by qPCR (n = 3 independent experiments). **f,** CCK-8 growth assay measuring cell growth rate of K562 WT and CRISPR clones (n = 3 independent experiments). **a,b,** P-values were calculated using two-tailed Welch t-test. **a,b,e,f,** All data are represented as means ± SE.



Supplementary Figure 9

Uncropped western blots and gel images.

Black boxes denote cropped images that are presented in the manuscript.