**Noncoding RNA landscape of mutant KRAS-driven cellular transformation**

Roman E. Reggiardo,1\* Haley Halasz,2\* Sreelakshmi Velandi Maroli,2 David Carrillo,2 Erin LaMontagne,1 Lila Whitehead,1 Eejung Kim,3,4 Shivani Malik,5 Jingchun Zhu,6 Jason Fernandes,1 Georgi Marinov,7 Eric Collisson,5 Angela Brooks,1,6,8,10

Daniel H. Kim1,6,8,9,10**\*\***

1Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, USA.

2Department of Molecular, Cell and Developmental Biology, University of California Santa Cruz, Santa Cruz, CA 95064, USA.

3Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.

4Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA.

5Department of Medicine, University of California San Francisco, San Francisco, CA 94158, USA.

6Genomics Institute, University of California Santa Cruz, Santa Cruz, CA 95064, USA.

7Department of Genetics, Stanford University School of Medicine, Palo Alto, CA 94305, USA.

8Center for Molecular Biology of RNA, University of California Santa Cruz, Santa Cruz, CA 95064 USA.

9Institute for the Biology of Stem Cells, University of California Santa Cruz, Santa Cruz, CA 95064 USA.

10California Institute for Quantitative Biosciences, University of California Santa Cruz, Santa Cruz, CA 95064 USA.

\*These authors contributed equally to this work.

**\*\*Corresponding author. Email: daniel.kim@ucsc.edu**

**ABSTRACT**

RAS genes are the most frequently mutated oncogenes in human cancer. However, the effects of RAS signaling on the vast noncoding transcriptome are unclear. We analyzed the transcriptomes of human lung and kidney cells transformed with mutant KRAS to define the landscape of RAS-regulated noncoding RNAs. We found that oncogenic RAS signaling upregulates noncoding transcripts throughout the genome that are enriched for transposable elements. These repetitive sequences are preferential targets of KRAB zinc-finger proteins, which are broadly downregulated in both mutant KRAS cells and lung adenocarcinomas. KRAS-mediated reprogramming of repetitive RNA induces an interferon response that contributes to cellular transformation. Our results reveal the extent to which mutant KRAS remodels the noncoding transcriptome, expanding the scope of genomic elements regulated by this fundamental signaling pathway.

Most of the human genome is noncoding and transcribed into RNA (*1, 2*), but how the noncoding transcriptome contributes to cancer formation is poorly understood. About half of the human genome is comprised of transposable elements (TE) (*3*), whose expression patterns are often altered in cancer (*4*). Moreover, TEs contribute substantially to the noncoding transcriptome and are present in the exonic sequences of thousands of long noncoding RNAs (lncRNAs) and other classes of regulatory RNAs (*5*). Noncoding RNA networks become disrupted in cancer (*6, 7*) and epigenetic reprogramming, where early activation of RAS signaling leads to coordinate regulation of noncoding RNAs in single cells (*8*). While RAS genes are among the most frequently mutated oncogenes in cancer (*9*), the extent to which RAS regulates the noncoding transcriptome during cellular transformation remains unknown.

To determine the landscape of noncoding RNAs affected by oncogenic RAS signaling, we performed RNA sequencing (RNA-seq) on human lung epithelial cells (AALE) that undergo malignant transformation upon introduction of mutant KRAS (*10*). We compared the transcriptomes of AALE cells transduced with control vector to AALEs that were transformed by mutant KRAS and analyzed the distribution of differentially expressed transcripts across the genome. Hundreds of lncRNAs were upregulated (n=279) or downregulated (n=409) by oncogenic RAS signaling, as well as many protein-coding mRNAs (n=4323 up, n=4711 down) and transcripts with retained introns (n=165 up, n=195) (Fig. 1A), revealing the broad extent to which mutant KRAS reprograms the coding and noncoding transcriptome. Compared to transcripts that were expressed but unchanged in the mutant KRAS versus control AALEs, a larger proportion of upregulated or downregulated lncRNAs and protein-coding mRNAs were comprised of TE sequences, while upregulated intron-retaining transcripts were also enriched for TEs (Fig. 1B), suggesting that TE sequence-containing loci in the genome are preferentially misregulated during malignant transformation.

To explore the biological pathways that are perturbed oncogenic RAS signaling, we performed gene set enrichment analysis (GSEA) (*11*) using genes that were differentially expressed in our mutant KRAS AALE cells. GSEA revealed that the most significantly enriched pathway was the interferon (IFN) alpha response, while the third most enriched pathway was IFN gamma response (Fig. 1C). These results suggest that the upregulation of TE sequence-containing RNAs by mutant KRAS activates an innate immune response in transformed AALEs.

We then investigated whether this mutant RAS-mediated IFN response was specific to lung cells or if unrelated cell types responded similarly. We performed RNA-seq on human embryonic kidney cells (HA1E-M) that were primed for oncogenic RAS-driven transformation (*12*) and analyzed how mutant KRAS altered their transcriptomes. We also observed that hundreds of lncRNAs were upregulated (n=165) or downregulated (n=223), along with protein-coding mRNAs (n=2635 up, n=2639 down) and retained-intron transcripts (n=119 up, n=237 down) (Fig. 2A), similar to what we found using mutant KRAS AALE cells. Moreover, differentially expressed RNAs were again enriched for TE sequences (Fig. 2B). When we performed GSEA, however, there was no enrichment for any IFN pathways in mutant KRAS-transformed HA1E-M cells, even though they were most significantly enriched for upregulated KRAS signaling (Fig. 2C). We found that both IFN gamma and IFN alpha response pathways were among the most significantly decreased gene sets (Fig. 2C), highlighting the tissue-specific differences in how the transcriptome is remodeled by mutant KRAS.

To investigate the molecular basis for IFN pathway activation in mutant KRAS AALE cells, we compared the expression patterns of differentially expressed IFN-stimulated genes in transformed AALEs and HA1E-M cells. AALEs with oncogenic RAS signaling upregulated the expression of the pattern recognition receptors (PRR) and cytosolic RNA sensors RIG-I and MDA5 (Fig. 3A) (*13*), while mutant KRAS HA1E-M cells showed no significant changes in their expression (Fig. 3B). These results suggest that TE sequence-containing RNAs induce a state of viral mimicry in KRAS-transformed AALE cells but not in KRAS-transformed HA1E-M cells.

We next performed knockdown studies of RIG-I and MDA5 in mutant KRAS AALE cells to determine the functional significance of PRR upregulation in the context of RAS-driven cellular transformation. RNA interference-mediated knockdown of KRAS resulted in significant loss of cell viability, while knockdown of MDA5 and RIG-I also significantly reduced viability (Fig. 3C), revealing the requirement for heightened levels of RIG-I and MDA5 expression in transformed AALE cells. To further characterize the nature of the IFN response in mutant KRAS AALEs, we performed single-cell RNA-seq (scRNA-seq) (n=1503 cells), which revealed that the IFN alpha gene signature was heterogeneous in KRAS-transformed AALEs, with a handful of individual cells exhibiting very high expression levels of the IFN alpha gene signature (Fig. 3D). We then analyzed the scRNA-seq data using a RIG-I/MDA5 induction gene signature, which showed that most of the individual cells within this population displayed high expression levels of this PRR signature (Fig. 3E).

To elucidate which TE classes might be involved in stimulating IFN-related gene expression, we analyzed the clusters in our scRNA-seq data (Fig. 4A) to determine the correlations between TE expression and various IFN gene signatures (Fig. 4B) (*14*). Alu elements were the most highly correlated TE class with the RIG-I/MDA5 induction gene signature (Fig. 4B), with the majority of Alu elements expressed in individual mutant KRAS AALE cells being AluSx1 elements (Fig. 4C). Given the known roles of KRAB zinc-finger proteins (KZNFs) in TE silencing, we examined whether Alu upregulation via oncogenic RAS signaling might result from the silencing of specific KZNFs. When we examined the differential expression of KZNFs in mutant KRAS AALEs, we observed a broad and significant downregulation of KRAB domain-containing and other zinc-finger proteins (Fig. 4D). In the mutant KRAS HA1E-M cells, however, no KZNFs were differentially expressed (Fig. 4E). Moreover, we examined KZNF chromatin immunoprecipitation sequencing (ChIP-seq) data (*15*) using the UCSC Repeat Browser and found that several of the significantly downregulated KZNFs in mutant KRAS AALEs bind to the consensus AluSx1 sequence (Fig. 4F). This suggests that suppression of these KZNFs via oncogenic RAS signaling leads to derepression of AluSx1 and other TEs during cellular transformation. This model is supported by broad and significant downregulation of KNZFs in mutant KRAS-driven lung adenocarcinomas when compared to healthy lung (Fig. 4G).

Collectively, our findings illustrate the impact of oncogenic RAS signaling on the noncoding transcriptome, while revealing tissue-specific differences in how transformed cells respond to TE sequence-containing RNAs. These conclusions are based on deeply sequencing and analyzing the transcriptomes of mutant KRAS-transformed cells at both the population and single-cell levels, building on previous work identifying noncoding RNAs that are coordinately regulated with RAS signaling genes in individual cells (*8*). The molecular basis for the state of viral mimicry we observe when only introducing mutant KRAS into AALE cells is different from similar TE-induced innate immune responses when cancer cells are treated with DNA methyltransferase inhibitors (*16, 17*), as we observe a prominent role for KZNFs in our system. Further studies will be required to test the functional consequences of upregulating hundreds of noncoding RNAs via oncogenic RAS signaling, as well as their potential utility as tissue-specific biomarkers of RAS-driven cancers.

**REFERENCES**

1. J. T. Lee, Epigenetic regulation by long noncoding RNAs. *Science* **338**, 1435-1439 (2012).

2. M. Kellis *et al.*, Defining functional DNA elements in the human genome. *Proc Natl Acad Sci U S A* **111**, 6131-6138 (2014).

3. E. S. Lander *et al.*, Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

4. K. H. Burns, Transposable elements in cancer. *Nat Rev Cancer* **17**, 415-424 (2017).

5. G. Bourque *et al.*, Ten things you should know about transposable elements. *Genome Biol* **19**, 199 (2018).

6. E. Anastasiadou, L. S. Jacob, F. J. Slack, Non-coding RNA networks in cancer. *Nat Rev Cancer* **18**, 5-18 (2018).

7. J. R. Evans, F. Y. Feng, A. M. Chinnaiyan, The bright side of dark matter: lncRNAs in cancer. *J Clin Invest* **126**, 2775-2782 (2016).

8. D. H. Kim *et al.*, Single-cell transcriptome analysis reveals dynamic changes in lncRNA expression during reprogramming. *Cell Stem Cell* **16**, 88-101 (2015).

9. B. Papke, C. J. Der, Drugging RAS: Know the enemy. *Science* **355**, 1158-1163 (2017).

10. A. S. Lundberg *et al.*, Immortalization and transformation of primary human airway epithelial cells by gene transfer. *Oncogene* **21**, 4577-4586 (2002).

11. R. K. Powers, A. Goodspeed, H. Pielke-Lombardo, A. C. Tan, J. C. Costello, GSEA-InContext: identifying novel and common patterns in expression experiments. *Bioinformatics* **34**, i555-i564 (2018).

12. E. Kim *et al.*, Systematic Functional Interrogation of Rare Cancer Variants Identifies Oncogenic Alleles. *Cancer Discov* **6**, 714-726 (2016).

13. A. J. Minn, Interferons and the Immunogenic Effects of Cancer Therapy. *Trends Immunol* **36**, 725-737 (2015).

14. J. L. Benci *et al.*, Opposing Functions of Interferon Coordinate Adaptive and Innate Immune Responses to Cancer Immune Checkpoint Blockade. *Cell* **178**, 933-948 e914 (2019).

15. M. Imbeault, P. Y. Helleboid, D. Trono, KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**, 550-554 (2017).

16. D. Roulois *et al.*, DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell* **162**, 961-973 (2015).

17. K. B. Chiappinelli *et al.*, Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* **162**, 974-986 (2015).

**ACKNOWLEDGEMENTS**

We thank members of the Kim Lab, Haussler Lab, Brooks Lab, and Carpenter Lab for helpful discussions. **Funding:** This work was supported by funds from the Baskin School of Engineering and the Ken and Glory Levy Fund for RNA Biology to D.H.K. R.E.R and H.H. were supported by the NHGRI-funded UCSC Genomic Sciences Graduate Training Program (NIH T32 HG008345). D.C. was supported by the NIGMS-funded UCSC IMSD Program (NIH R25 GM058903). **Author Contributions:** D.H.K. conceived and designed the study and wrote the manuscript, R.E.R. performed computational analysis and generated figures, H.H., S.V.M., D.C., E.L., L.W., E.K., and S.M. performed experiments, J.Z., J.F., E.C., and A.B. provided resources, and G.M. performed computational analysis. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All RNA sequencing data have been deposited in NCBI GEO (GSE120566).

**MATERIALS AND METHODS**

Cell Lines

The AALE stable cell lines pBABE-mCherry Puro (control) and pBABE-FLAG-KRAS(G12) Zeo (mutant KRAS) were generated using retroviral transduction, followed by selection in puromycin of zeocin, respectively, 2 days post-infection. Both lines were cultured in SABM Basal Medium (Lonza SABM basal medium) with added supplements and growth factors (Lonza SAGM SingleQuot Kit Suppl. & Growth Factors). AALE cell lines were maintained using Lonza’s Reagent Pack subculture reagents. The HA1E cell lines were generated using lentiviral transduction (pLX317) to generate control and mutant HA1E pLX317-KRAS(G12) stable cell lines using puromycin selection, and cells

were cultured in MEM-alpha (Invitrogen) with 10% FBS (Sigma) and 1% penicillin/streptomycin (Gibco). All cell lines tested negative for mycoplasma.

siRNA Knockdowns

AALEs were seeded at 1x106 cells per well of a 6-well plate in complete growth medium, then reverse transfected with 30pmol siRNA using RNAiMAX lipofectamine according to manufacturer’s protocol. Cells were grown for 3 days in transfection medium under standard culture conditions and then harvested for RNA isolation and qPCR as previously described.

Cell Viability Assay

2x104 cells were subtracted from each siRNA transfection well at the time of transfection and seeded into individual wells of an ultra-low adhesion 96-well plate. The cells were grown in standard culture conditions for 4 days. They were then harvested, and ATP production was measured using the Cell TiterGLO Luminescent Cell Viability Assay (Promega) following the manufacturer’s protocol. Luminescence was measured on a Perkin Elmer VICTOR light 1420 Luminescence Counter.

RNA Isolation & Purification

For AALE cell lines, bulk RNA was isolated from cells using Quick-RNA MiniPrep kit (Zymogen). All RNA was quantified via NanoDrop-8000 Spectrophotometer. For HA1E cell lines, bulk RNA was isolated using RNeasy Mini Kit (Qiagen) and quantified via Qubit RNA BR assay kit (Thermo).

qPCR

cDNA was transcribed from 1ug RNA using iScript cDNA Synthesis Kit (Bio-Rad) according to manufacturer protocol. cDNA was diluted 1:6 and run with iTaq Universal SYBR Green Supermix (Bio-Rad) on ViiA 7 Real-Time PCR System according to manufacturer protocol. Cycle Threshold (CT) values were converted using Standard analysis. Values obtained for target genes were normalized to HPRT.

Library Preparation for Bulk RNAseq

For AALE cell lines, 1ug of total RNA was used as input for the TruSeq Stranded mRNA Sample Prep Kit (Illumina) according to manufacturer protocol. Library quality was determined through the High Sensitivity DNA Kit on a Bioanalyzer 2100 (Agilent Technologies). Multiplexed libraries were sequenced as HiSeq400 100PE runs. For HA1E cell lines, 1ug of total RNA was used for mRNA enrichment with Dynabeads mRNA DIRECT kit (Thermo). First strand cDNA was generated with AffinityScript Multiple Temperature reverse transcriptase with oligo dT primers. Second strand cDNA was generated with mRNA Second Strand Sythesis Module (New England Biolab). DNA was cleaned up with Agencourt AMPure XP beads twice. Qubit dsDNA High Sensitivity Assay was used for concentration measurement. 1ng of dsDNA was further subjected to library preparation with Nextera XT DNA sample prep kit (Illumina) per manufacturer instructions. Library size distribution was confirmed with Bioanalyzer (Agilent). Multiplexed libraries were sequenced as NextSeq500 75PE runs.

Library Preparation for Single Cell RNAseq

For single cell RNAseq, 1x106 cells were harvested and re-suspended in 1mL 1X PBS/0.04% BSA (1000 cells/ul) according to the cell preparation guidelines in the 10X Genomics Chromium Single Cell 3’ Reagent Kit User Guide. GEMs were generated from an input of 3,500 cells. We used the 10X Genomics Chromium Single Cell 3’ Reagent Kits version 2 for both the GEM generation and subsequent library preparation and followed the manufacturer’s reagent kit protocol. Quantification of all RNAseq libraries was performed by QB3 at UC Berkeley. RNAseq libraries were sequenced as HiSeq4000 100PE runs.

Statistical Analysis

All quantitative data for functional assays has been reported as means ± standard deviation. Statistical significance for these was calculated using a t-test and p-values <0.05 were considered significant.

*All sub-sections below are associated with R scripts hosted on* [*GitHub*](https://github.com/rreggiar/KRAS-AALE-Paper/blob/master/KRAS-AALE-CODE/) *and organized in the “aale.2019.analysis.notebook.Rmd”*

RNA-seq psuedoalignment and quantification

All *fastq* files were trimmed with *Trimmomatic 2 (0.38)* using the Illumina NextSeq PE adapters. The resulting trimmed files were assessed with *FastQC* and then passed through the following analytical pipeline:

Salmon (0.14.1): pseudoalignment of RNA-seq reads performed with *Salmon* using the following arguments:

--validateMappings --rangeFactorizationBins 4 --gcBias --numBootstraps 10

with the ouput directories being processed in the following manners.

Sleuth: transcript differential expression was performed using *Sleuth*. Upon completion, the transcripts with q-values below 0.05 in the likelihood-ratio test were used to filter salmon output from which log2fc was manually calculated and paired to the sleuth output.

DESeq2: *Salmon* output was imported to DESeq object using *tximport* and differential expression analysis was performed with standard arguments.

Transposable Element Content Analysis

5’/3’ UTR: mRNA sequences for all relevant genes were retrieved from the UCSC Genome Browser using the Table Browser functionality. Promoter sequences were defined as all genomic content 2000 nucleotides upstream of the transcription start site. FASTA files containing either mRNA or promoter sequences were cleared of duplicates and renamed using a custom *Python* script. All resulting FASTA files were processed using *RepeatMasker* with the following option flags:

*-no\_is -nolow -s -species Human -pa 8*

Parsing of *RepeatMasker* output was performed with bash command line tools and custom *Python* scripts. *SalmonTE* with default settings was used to quantify Transposable Element transcript expression from aggregate bulk RNA sequencing data and output was parsed using *R*. Analysis of all data was performed and visualized in *R* using custom scripts implementing the *Tidyverse* package.

Exon Overlap: a whole genome *.gtf* file was downloaded from the UCSC genome browser Table browser utility. This file was parsed and merged with the GENCODE v.29 reference transcriptome. This modified *.gtf* (now a *.bed* file) was passed to *bedtools* where the overlap function was used with the following arguments:

-a modified.gtf.bed -b all.ucsc.rmsk.genes.bed -wao -s > retained.overlap.bed

alongside a whole genome *.gtf* retrieved as described above except generated from the repeat-masked browser track. The resulting overlapped bed file was processed and visualized using custom *R* scripts.

Zinc Finger Protein Analysis

ChIP-exo data and supplementary information were extracted from supplementary data provided by *Imbeault et al*. ZNF genes were cross referenced with *DESeq2* and *RepeatMasker* outputs to extract relevant differential expression data of ZNF proteins and Transposable Element transcripts using *R*. *RepeatMasker* output from promoter analyses was cross referenced with ChIP-exo target data to identify potential regulatory targets of differentially expressed KZNFs. Only KZNF targets with ‘score’ [see Imbeault *et al*] >= 75 were kept for analysis. Analysis of all data was performed and visualized in *R* using custom scripts.

Gene Set Enrichment Analysis

Genes determined to be significantly differentially expressed in *DESeq2* output were first ‘pre-ranked’ in *R* by the following metric:

Score metric = sin(log2FoldChange) \* -log10(p-value)

The resulting ranked files objects were processed using the *R* package *fgsea* alongside gene set files downloaded from msigdb using the *R* package *msigdbr*. Additional code was written for select visualizations.

Gene Ontology Analysis

Upregulated gene names were extracted from *DESeq2* output using bash command line tools. Name lists were pasted into the *Gene Ontology Consortium*’s *Enrichment Analysis* tool powered by *PANTHER*. Output data was exported as *.txt* files and parsed using bash command line tools. Parsed data was visualized using custom *R* scripts.

Single Cell Analysis:

*All single cell analyses are associated with R scripts hosted on* [*GitHub*](https://github.com/rreggiar/KRAS-AALE-Paper/blob/master/KRAS-AALE-CODE/) *and organized in the “aale.2019.single.cell.analysis.Rmd”*

10x Processing: Single cell output data was processed using 10x pipeline *CellRanger* [The *mkfastq* functionality was used to generate *fastq* files for further downstream analysis. Output was also aggregated and quantified using the *aggr* and *count* functionalities, respectively. This output was visualized using the 10x Loupe browser.

Downstream Analysis: *fastq* files generated above were passed to Salmn *alevin* with the following arguments:

--libtype A –chromium –dumpCsvCounts -p 16.

*alevin* was used to psuedoalign the libraries to both the GENCODE v.29 reference transcriptome as well as a composite transcriptome reference built by combining the GENCODE v.29 reference with one built from the GRCh38\_rmsk\_TE.gtf hosted by the Hammel lab at <http://labshare.cshl.edu/shares/mhammelllab/www-data/TEToolkit/TE_GTF/> . A salmon index was built from this reference with standard arguments. These alevin output matrices were imported into R using *tximport*. GSEA/ cluster correlations were calculated using the *R corr()* function. Normalization and clustering were performed with *Seurat* and additional code was written to handle select visualizations.

TCGA ZNF analysis:

TCGA-LUAD phenotype and normalized count data were downloaded from the UCSC Xena browser data repository. The files were combined and patients were grouped by their KRAS mutation status and identity. These data were compared to and visualized alongside of data generated from our analysis using custom *R* code.

**FIGURE LEGENDS**

**Fig. 1. Lung cell-specific transcriptome reprogramming by mutant KRAS.** (**A**) Chromosome-level distribution of differentially expressed RNAs in mutant lung epithelial cells. Shown are the three most abundant biotypes from RNA-seq data. (**B**) Proportion of exons that overlap a transposable element (TE) for all genes detected and differentially expressed in mutant lung epithelial cells, separated by biotype. (**C**) Gene set enrichment analysis (GSEA) pathways sorted by normalized enrichment score (NES) in mutant lung epithelial cells.

**Fig. 2.** **Kidney cell-specific transcriptome reprogramming by mutant KRAS.** (**A**)

Chromosome-level distribution of differentially expressed RNAs in mutant kidney cells. Shown are the three most abundant biotypes from RNA-seq data. (**B**) Proportion of exons that overlap a transposable element (TE) for all genes detected and differentially expressed in mutant kidney cells, separated by biotype. (**C**) Gene set enrichment analysis (GSEA) pathways sorted by normalized enrichment score (NES) in mutant kidney cells.

**Fig. 3. Interferon-stimulated gene expression heterogeneity in transformed cells.** Differentially expressed interferon-stimulated genes in (**A**) mutant lung epithelial cells and (**B**) mutant kidney cells. (**C**) Cell viability in mutant lung epithelial cells transfected with indicated small interfering RNAs. (**D** and **E**) Uniform manifold approximation and projection (UMAP) visualization of single-cell RNA-seq data from mutant lung epithelial cells showing expression of indicated metagenes.

**Fig. 4. Broad suppression of KRAB zinc finger proteins in lung cancer cells.** (**A**) UMAP visualization of single-cell RNA-seq clustering of mutant lung epithelial cells. (**B**) Correlation between different classes of transposable elements (TEs) and indicated metagenes in single-cell RNA-seq data. (**C**) Correlation plot between ALU and AluSx1 in single-cell RNA-seq data. Differentially expressed zinc finger proteins in (**D**) mutant lung epithelial cells and (**E**) mutant kidney cells. (**F**) ChIP-seq data from indicated zinc finger proteins showing binding to AluSx1. (**G**) Significantly repressed zinc finger proteins in mutant KRAS lung adenocarcinomas compared to normal lung samples.