***Science* Supplementary Materials Template Instructions**

This is the *Science* template for presenting and formatting your supplementary materials. To organize your supplementary materials section, please follow the instructions below. Once formatted, you should delete this first page of instructions.

**Overview:**

Supplementary materials present additional information in support of the conclusions of your paper, such as a description of the materials and methods, controls, or tabulated data presented in Tables or Figures. It will generally consist of one integrated PDF file. Audio or movie files or large data Tables can be presented as separate files. Further information is available at: <http://www.sciencemag.org/authors/instructions-preparing-initial-manuscript#format-supplemental>.

This section should not be used for additional discussion, analysis, or interpretations. It is not to be used as a forum to critique other publications.

References can be cited in the supplementary text section. These should be cited in order following the references in the main text as per *Science* style (i.e., italicized number in parentheses). One full reference list should be provided for all cited references at the end of the main paper.

**Using the Template**

Paste the title, author list, and corresponding author email address(es) from the main text file onto the cover page. On the cover page, complete the relevant description of the SM and delete text that does not apply.

Copy and paste relevant text into each appropriate section of the template.

Each Figure or Table should be on a separate page and can be placed above each caption. To add additional captions, simply copy and paste (repeatedly) the last caption template. Large Tables that extend beyond the width of the page should be provided as separate files in an appropriate spreadsheet format (.xlsx or similar).

Large amounts of text can be grouped by subheads. To repeat subheads, simply copy the subhead and repeat/rename.



Supplementary Materials for

Paste Your Title Here—No Paragraph Breaks

Paste the full author list here, no paragraph breaks.

Correspondence to: [xxxxx@xxxx.xxx](mailto:xxxxx@xxxx.xxx)

**This PDF file includes:**

Materials and Methods

Supplementary Text

Figs. S1 to Sx

Tables S1 to Sx

Captions for Movies S1 to Sx

Captions for Audio S1 to Sx

Captions for Data S1 to Sx

**Other Supplementary Materials for this manuscript include the following:**

Movies S1 to Sx

Audio S1 to Sx

Data S1 to Sx [paste data table titles in a list]

Materials and Methods

Experimental Design

To investigate how oncogenic KRAS impacts the transcriptome of lung cancer cells, we utilized an *in vitro* model in which we transiently transfected AALE cells with an integrating KRAS G12V plasmid. We then performed both bulk and single-cell RNA-sequencing during different stages of the transformation process. We obtained AALEs with a stably integrated KRAS G12D, which was also included in RNA-seq experiments. To test the functional significance of observations made from the RNA-seq data, we performed loss of function studies in the KRAS G12D cells using RNA interference, chemical inhibition and cell growth assays. All experiments were done in biological triplicates.

Cell Lines

The wild-type female AALE cell line was provided by Scott Randell’s Lab at University of North Carolina. The female AALE stable cell lines pBABE-mCherry Puro and pBABE-FLAG-KRAS(G12) Zeo were provided by Eric Collisson’s Lab at University of California, San Francisco. 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) for complete growth medium. AALE cell lines were maintained using Lonza’s Reagent Pack subculture reagents and standard tissue culture procedures. The HA1E cells were derived from female human embryonic kidney cells in Bill Hahn’s Lab at Dana Farber Cancer Institute. Wild-type HA1E cells and the HA1E pLX317-KRAS(G12) stable cell line were cultured in MEM-alpha (Invitrogen) with 10% FBS (Sigma) and 1% penicillin/streptomycin (Gibco). All cell lines tested negative for mycoplasma.

*In vitro* KRAS-driven Transformation

100mm dishes of 80% confluent AALE cells were transfected with 12ug of pBABE-KRAS(G12) Puro plasmid from Addgene (#9052) using Invitrogen Lipofectamine 2000 DNA transfection reagent and protocol. After 24hrs, the medium was replaced with SAGM supplemented with 2ug/ml of puromycin. Transfected cells were kept under puromycin selection for 5, 14, and 21 days. Cells were trypsinized and harvested by centrifugation at 300 g x 3 minutes for RNA extraction at each time point respectively.

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. Refer to supplementary table for complete list of siRNAs.

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 manufacturers protocol. Luminescence was measured on a Perkin Elmer VICTOR light 1420 Luminescence Counter.

DNMT Inhibitor Treatment

AALEs were seeded in 6 well plates at 2.5x105 cells per well one day prior to treatment. The cells were treated with 3nM 5-Aza-2’-Deoxicytidine (Millipore #189826) for 24 hours. After 24 hours, cells were harvested, RNA was isolated and qPCR was performed as described. 1.5x103 cells were seeded into wells of an ultra-low adhesion 96-well plate and treated with 3nM 5-Aza-2’-Deoxicytidine for 24 hours. Cells were grown for 3 days and then ATP production was measured as previously described.

RNA Isolation & Purification

For AALE cell lines, bulk RNA was isolated from cells using Quick-RNA MiniPrep kit (Zymogen). For DNMT inhibitor and siRNA experiments RNA was isolated using Zymogen’s Direct-zol Miniprep Plus kit. 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 Supermix 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). 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. All AALE cell line 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 analyses were conducted in Microsoft Excel version 16.18.

Transposable Element Content Analysis

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.

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. To determine genome distribution of ZNF genes, the UCSC genome browser Table Browser functionality was used to extract a BED12 file with all exonic coordinates corresponding to ZNFs sourced from. Analysis of all data was performed and visualized in *R* using custom scripts implementing the *Tidyverse* package.

Gene Set Enrichment Analysis

*DESeq2* output was first ‘pre-ranked’ by *log2FoldChange* in *R* and then loaded into the *javaGSEA* desktop application. The *GSEA-Preranked* functionality with default settings was used along with all *MSigDB* Hallmark gene sets to determine enrichment. *.xls* output was parsed and visualized in *R* using scripts implementing the *Tidyverse* package.

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 Regression Analysis

Demultiplexed 10x chromium FASTQ files were aligned to both *RepeatMasker* consensus annotation for hg38 as well as gencode.v29.transcript.fa using *salmon alevin* with options:

*--libType A –chromium –index \*index reference with salmon index\* -p 10 –tgMap \* transcript to*

*Gene name tsv \* --dumpCsvCounts .* The cell-count matrix was assembled from the row, column, and matrix files produced in output using command line tools and *R*. Gene columns in matrices were filtered to only include genes that were counted at least n = number of cells times across the entire library. Simple regression analysis was performed using the *R* function ‘lm’.

Supplementary Text

Subhead

Type or paste text here. This should be additional explanatory text, such as: extended technical descriptions of results, full details of mathematical models, extended lists of acknowledgments, etc. It should not be additional discussion, analysis, interpretation, or critique.

<insert page break then Fig. S1 here>

Fig. S1.

Type or paste caption here. Create a page break and paste in the Figure above the caption.

<insert page break then Fig. S2 here>

Fig. S2.

Type or paste caption here. Create a page break and paste in the Figure above the caption.

<insert page break then Fig. S3 here>

Fig. S3.

Type or paste caption here. Create a page break and paste in the Figure above the caption.

<insert page break here>

Table S1.

Primers and siRNA sequences.

|  |  |  |
| --- | --- | --- |
| **Type** | **Target** | **Sequence** |
| Primers | HPRT | FWD 5’-GCAGCCTAGGTCTCTGG-3’  REV 5’-GTTCCTTTTCACCAGCAAGCT-3’ |
| Primers | KRAS | FWD 5’-GGACTGGGGAGGGCTTTC-3’  REV 5’- GCCTGTTTGTGTCTACTGTTCT-3’ |
| Primers | MDA5 | FWD 5’-GAGGAATCAGCACGAGGAATAA-3’  REV 5’-TCAGATGGTGGGCTTTGAC-3’ |
| Primers | RIG-I | FWD 5’-AAGAGCAAGAGGTAGCAAGTG-3’  REV 5’-CCCATGTCTGAAGGCGTAAA-3’ |
| Primers | OAS2 | FWD 5’-CTGGAGCTGGTCACACAATATC-3’  REV 5’-GAAACTTCCTCACGGTCTCATC-3’ |
| DsiRNA | KRAS | 5’-ACAGGAAGCAAGUAGUAAUUGAUGG-3’  3’-GAUGUCCUUCGUUCAUCAUUAACUACC-5’ |
| DsiRNA | MDA5 | 5’-GUCAUCAACACCAACAAAGAAGCAGT-3’  3’-UACAGUAGUGUGGUUGUUUCUUCGUCA-5’ |
| DsiRNA | RIG-I | 5’-CAGAAUCUUAGUGAGAAUUCAUGTC-3’  5’-CGGUCUUAGAAUCACUCUUAAGUACAG-3’ |
| DsiRNA | Non-targeting | 5’-CGUUAAUCGCGUAUAAUACGCGUAT-3’  3’-CAGCAAUUAGCGCAUAUU-5’ |

Table S2.

Type or paste caption here. Create a page break and paste in the Table above the caption.

<insert Table S2 here followed by a page break >

Movie S1.

Type or paste caption here.

Audio S1.

Type or paste caption here.

Data S1. (separate file)

Type or paste caption here.

ROMAN REFS:

Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-4.0*.  
2013-2015 <http://www.repeatmasker.org>.

SalmonTE -- Guo, C., Jeong, H.-H., Hsieh, Y.-C., Klein, H.-U., Bennett, D.A., Jager, P.L.D., Liu, Z., and Shulman, J.M. (2018). Tau Activates Transposable Elements in Alzheimer’s Disease. Cell Reports *23*, 2874–2880.

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

R Core Team (2018). R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Hadley Wickham (2017). tidyverse: Easily Install and Load the 'Tidyverse'. R package

version 1.2.1. <https://CRAN.R-project.org/package=tidyverse>

GSEA -- Subramanian, Tamayo, et al. 2005 Proc Natl Acad Sci U S A 102(43):15545-50

MSigDB -- Liberzon et al. 2011 Bioinformatics 27(12):1739-40

ALL GO REFS:

Ashburner et al. Gene ontology: tool for the unification of biology (2000) *Nat Genet* **25(1)**:25-9. [Online at Nature Genetics](https://www.ncbi.nlm.nih.gov/pubmed/10802651).

GO Consortium, Nucleic Acids Res., 2017

Mi et al., Nucleic Acids Res., 2017