Genes & Diseases (2026) 13, 101840

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.keaipublishing.com/en/journals/genes-diseases

FULL LENGTH ARTICLE

# *In vivo* multiplexed modeling reveals diverse roles of the TBX2 subfamily and *Egr1* in *Kras*-driven lung adenocarcinoma

Athar Khalil [a,b,*], Trang Dinh [a], Meaghan Parks [a], Rebecca C. Obeng [c,d], Berkley Gryder [a,b], Adam Kresak [d], Yuxiang Wang [a], Jeff Maltas [a,b], Madeline Bedrock [a], Xiangzhen Wei [a], Zachary Faber [a], Mira Rahm [a], Jacob Scott [e], Thomas LaFramboise [a], Zhenghe Wang [a,b,*], Christopher McFarland [a,b]

[a] *Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA*
[b] *Cancer Genomics and Epigenomics Program, Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH 44106, USA*
[c] *Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA*
[d] *Department of Pathology, University Hospitals Cleveland Medical Center, Cleveland, OH 44106, USA*
[e] *Translational Hematology & Oncology Research, Cleveland Clinic, Cleveland, OH 44106, USA*

Received 3 June 2025; accepted 14 July 2025
Available online 3 September 2025

**KEYWORDS**
Early growth response 1;
*In vivo* models;
TBX2 subfamily;
Tuba-seq;
Tumor suppressors

**Abstract** The TBX2 subfamily of T-box transcription factors (*e.g.*, *Tbx2*, *Tbx3*, *Tbx4*, *Tbx5*) plays an essential role in lung development. Down-regulation of these genes in human lung adenocarcinoma suggests that these genes may be tumor-suppressive; however, because down-regulation appears to occur primarily via epigenetic change, it remains unclear if these changes causally drive tumor progression or are merely the consequence of upstream events. Herein, we developed the first multiplexed mouse model to study the impact of TBX2 subfamily loss, alongside associated signaling genes (*Egr1*, *Chd2*, *Tnfaip3a*, and *Atf3*) in *Ras*-driven lung cancer. Using tumor-barcoding with high-throughput barcode sequencing (TuBa-seq), a high-throughput tumor-barcoding system, we quantified the growth effects of these knockouts during early and late tumorigenesis. *Chd2* knockout suppressed both tumor initiation and progression, whereas *Tnfaip3* knockout enhanced

* Corresponding authors. Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA.
*E-mail addresses:* axk1472@case.edu (A. Khalil), zxw22@case.edu (Z. Wang).
Peer review under the responsibility of Chongqing Medical University.

https://doi.org/10.1016/j.gendis.2025.101840

tumor initiation and overall tumor growth. *Tbx2* loss showed stage-specific effects on tumor development. Notably, *Egr1* emerged as a strong tumor suppressor and its knockout resulted in approximately a fivefold increase in tumor size at 20 weeks (two-sample $t$-test, $p < 0.05$), exceeding the impact observed with *Rb1* loss. Transcriptomic analyses of *Egr1*-deficient tumors suggested immune dysregulation, including heightened inflammation and potential markers of T cell exhaustion in the tumor microenvironment. These findings indicate that *Egr1* may play a role in suppressing tumor growth through modulating immune dynamics, offering new insights into the interplay between tumor progression and immune regulation in lung adenocarcinoma.


## Introduction

Lung adenocarcinoma (LUAD) is a genetically diverse cancer driven by numerous oncogenic and tumor-suppressive events.[1] The T-box gene family, comprising eighteen members in mammals, encodes transcription factors that are crucial for embryonic development and tissue homeostasis.[2,3] The four members of the TBX2 subfamily (*Tbx2*, *Tbx3*, *Tbx4*, and *Tbx5*) play an indispensable role in lung development, with their expression being crucial for various aspects of normal lung morphogenesis and differentiation.[4–7] In previous studies, we observed marked down-regulation of the TBX2 subfamily genes in clinical LUAD specimens. The down-regulation of these genes appears to be predominantly mediated by epigenetic mechanisms, particularly DNA methylation dysregulation.[8,9] Because these clinical analyses are inherently correlative, it remains unclear whether the down-regulation of these genes causally drives tumor progression or is merely a consequence of broader transcriptional reprogramming. Reactivation of these genes in LUAD cell lines led to reduced cell growth, suggesting a potential tumor suppressive function.[10] This effect is amplified in the presence of a Kirsten rat sarcoma viral oncogene homolog (*KRAS* ) mutation, the most common oncogenic event in human LUAD.[11] In other cancer types, TBX2 genes play a differential role in regulating cell cycle progression, proliferation, senescence, apoptosis, inflammation, and metastasis.[12] With such a diverse and central role in cancer, these genes were shown to function as either oncogenes or tumor suppressors depending on the cellular context (Fig. 1A).[13–15] Hence, faithful models of TBX2 gene loss in an *in vivo* lung environment are needed to understand the causal role of these genes in LUAD, particularly in the context of *Kras*-driven oncogenesis.

Herein, we developed the first multiplexed mouse model to study TBX2 subfamily loss along with potential downstream effector genes (*Egr1*, *Chd2*, *Tnfaip3a*, and *Atf3*) identified through our *in vitro* studies in *Ras*-driven lung cancer.[10] Our approach combines tumor-barcoding with high-throughput barcode sequencing (TuBa-seq).[16] Tumor initiation was carried out using pools of barcoded Lenti-sg*RNA*/Cre viruses, and the resulting tumors were analyzed at two distinct time points, thereby comprehensively assessing multiple stages of tumorigenesis, including tumor initiation, growth, and progression of Kras$^{G12D}$-driven lung tumors, across nine distinct genotypes. Notably, early growth response 1 (*Egr1*) emerged as a critical regulator, with its loss having a more pronounced impact on tumor progression than the loss of the well-established tumor suppressor gene retinoblastoma 1 (*Rb1*).

## Materials and methods

### Mice and tumor initiation

All animal experiments conducted in this study received approval from the Institutional Animal Care Facility at Case Western Reserve University, under protocol number 2020-0099. Kras$^{LSL-G12D;\ Rosa26CAG-LSL\text{-}Cas9-GFP}$ mice have been described.[17,18] Mice were on a mixed BL6/129 background. Approximately equal numbers of males and females were used for each experiment. The corresponding figure legends specify the number of mice used in each experiment. Lung tumor initiation was achieved through intratracheal administration of viral-Cre vectors, following the established protocol.[19] Ketamine/xylazine drugs were used for anesthesia before the intubation step, and atipamezole was used directly afterwards as a reversal drug. The assessment of tumor burden encompassed various methodologies, including fluorescence microscopy, lung weight measurements, and histological analyses, all of which are comprehensively detailed in the appropriate sections of this study.

### Design and generation of Lenti-sgRNA/Cre vectors

We designed lentiviral vectors to encompass Cre alongside single guide RNAs (sgRNAs) targeting specific genes under investigation (putative tumor suppressors and controls) in LUAD. These genes included *Tbx2*, *Tbx3*, *Tbx4*, *Tbx5*, *Egr1*, *Tnfaip3*, *Chd2*, and *Atf3* (Table S1). Additionally, vectors carrying inert guides sg*Neo1* and sg*Neo2* were generated as internal controls. For validation purposes, sg*Rb1* served as a positive control, while an sgRNA targeting proliferating cell nuclear antigen (*Pcna*), an essential gene, acted as a negative control. The sgRNA sequences were meticulously crafted using CRISPRpick. We identified all feasible 20-bp sgRNAs with an NGG protospacer-adjacent motif (PAM) targeting each tumor suppressor gene. These sgRNAs were evaluated for predicted on-target cutting efficiency

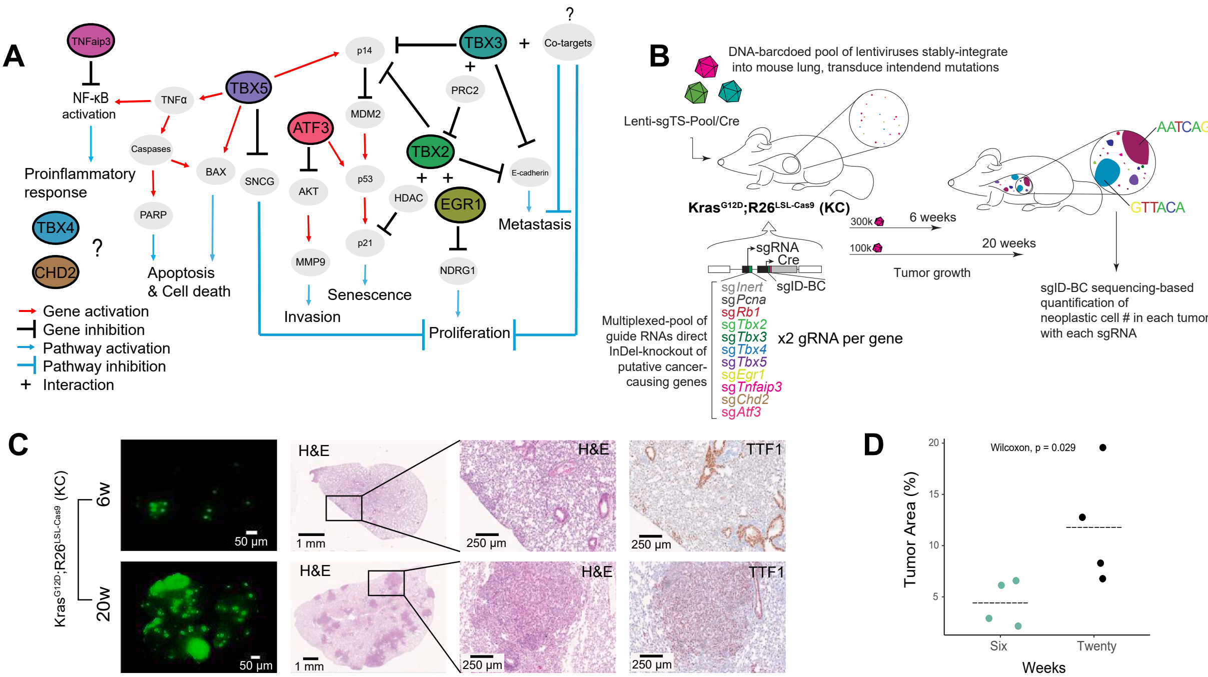


**Figure 1** TBX2 signaling in cancer and TuBa-seq approach for evaluating their role in *Kras*-driven lung adenocarcinoma. **(A)** Schematic representation of the molecular basis of TBX2-associated signaling pathways in cancer. *TBX2*, *TBX3*, and *TBX5* contribute to the dysregulation of multiple cancer hallmarks, including cell proliferation, senescence, apoptosis, invasion, metastasis, and the pro-inflammatory response. *EGR1*, *TNFaip3*, *ATF3*, *Tbx4*, and *Chd2* are all modulated by TBX2 signaling in lung cancer cell lines and recurrently dysregulated in human lung cancers. While *EGR1*, *TNFAIP3*, and *ATF3* all directly modulate key carcinogenic pathways, the mechanisms by which *TBX4* and *CHD2* influence cancer remain largely unexplored.[14,15,34,42–48] **(B)** Experimental schematic of the TuBa-seq approach used to investigate combinatorial inactivation of potential lung adenocarcinoma regulators *in vivo*. Tumor initiation was achieved through intratracheal intubation with Lenti-sgTS-Pool/Cre in Kras$^{LSL-G12D}$; Rosa26$^{CAG-LSL\text{-}Cas9-GFP}$ mice. The Lenti-sgTS-Pool contained two inert sgRNA vectors along with a positive control (sg*Rb1*) and a negative control (sg*Pcna*). Each of the eight genes studied was targeted using two distinct sgRNAs. Each sgRNA vector included a unique sgID and a random barcode, allowing quantification of individual tumor sizes via deep sequencing. The genotype, time points, and lentiviral titers are as indicated. **(C)** Representative images of lung lobes from mice at 6 weeks and 20 weeks after tumor initiation. Images include fluorescence dissecting scope views, hematoxylin-eosin-stained sections, and TTF1-stained (lung epithelial marker) sections of lung lobes. Scale bars are indicated on each image. **(D)** Quantification of the percent tumor area in representative mice revealed a significant increase at 20 weeks versus 6 weeks (Wilcoxon test, $p < 0.05$). Each dot represents an individual mouse, with horizontal bars indicating the mean tumor area.

through an sgRNA design/scoring algorithm. From this analysis, two unique sgRNAs per tumor suppressor gene were selected, prioritizing those with the highest predicted cutting efficiencies. Preference was given to sgRNAs targeting exons conserved across all known splice isoforms (ENSEMBL), positioned closest to splice acceptor or donor sites, located earliest in the gene-coding region, and situated upstream of annotated functional domains (InterPro; UniProt) (Table S1).

The Lenti-U6-sgRNA-sgID-barcode-Pgk-Cre vector was cloned in collaboration with Genewiz. Initially, the sgRNA sequence of the pLenti-sgRb1/Cre vector (Addgene #89647) was replaced by GCGAGGTATTACCGGCGTATCATCCGCG using site-directed mutagenesis, thereby creating pLenti-BaeI-Pgk-Cre. This replacement sequence incorporated a recognition site for the type IIS restriction endonuclease BaeI, facilitating swift replacement of the sgRNA sequence. Subsequently, forward and reverse single-stranded oligonucleotides containing the sgRNA sequence and complementary overhangs were annealed and ligated into the BaeI-linearized pLenti-BaeI-Pgk-Cre vector utilizing T4 DNA ligase. The barcode oligo primer, featuring the 8-nucleotide sgID sequence and 20-nucleotide degenerate barcode, was generated and ligated into the vectors according to established protocols.[16]

## Production, purification, and titering of lentivirus

To prevent barcode-sgRNA uncoupling due to lentiviral template switching during reverse transcription of the pseudo-diploid viral genome, individual barcoded Lenti-sgRNA/Cre vectors were generated separately.[20] HEK293T cells were cultured in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and transfected with each of our barcoded Lenti-sgRNA/Cre plasmids, along with pCMV-VSV-G (Addgene #8454) envelope plasmid and pCMV-dR8.2 dvpr (Addgene #8455) packaging plasmid, using polyethylenimine. Following transfection, the cells were treated with 20 mM sodium butyrate after 8 h, the culture medium was changed 24 h later, and supernatants were collected at 36 h and 48 h after transfection. Cell debris was removed using a 0.45 μm syringe filter unit (Millipore SLHP033RB), and each lentiviral vector was concentrated by ultracentrifugation at 25,000 rpm at 4 °C for 1.5 h. The concentrated virus was resuspended in phosphate-buffered saline and stored at −80 °C. Vector titers were determined by transducing Rosa26LSL-YFP mouse embryonic fibroblasts with 10 μL of each Lenti-sgRNA/Cre vector, assessing the percentage of YFP-positive cells via flow cytometry, and normalizing the titer to a lentiviral preparation with a known titer. At least 50 k cells were sorted per sample. The cell line was confirmed negative for mycoplasma contamination. Lentiviral vectors were thawed and pooled immediately before administration to mice.

## Genomic DNA isolation from mouse lungs

Genomic DNA extraction was carried out from bulk tumor-bearing lung tissue obtained from each mouse. Before tissue homogenization, six benchmark control cell lines, comprising approximately $1 \times 10^5$ cells for three cell lines and 100 cells for three cell lines, each carrying unique sgID-BCs, were introduced ("spiked-in") to every sample, as previously described.[36] This addition facilitated the subsequent calculation of the absolute number of neoplastic cells in each tumor based on the sgID-BC reads. The entire lung tissues or right lobe from each mouse, along with the benchmark cell lines, underwent homogenization using a gentleMACS dissociator. Homogenization was performed in 6 mL lysis buffer (100 mM NaCl, 20 mM Tris, 10 mM EDTA, 0.5% SDS) supplemented with 200 μL of 20 mg/mL proteinase K (Life Technologies, AM2544). The homogenized tissue was incubated at 55 °C overnight. Subsequently, DNA extraction was carried out through phenol—chloroform extraction and ethanol precipitation from approximately 1/6th of the total lung lysate using standard protocols. DNA concentrations were determined using a nanodrop.

## Library construction and sequencing of sgID-BC for quantitative analysis

Libraries were prepared through the amplification of the sgID-BC region from 32 μg of genomic DNA per mouse. The sgID-BC region of the integrated Lenti-sgRNA-BC/Cre vectors underwent PCR amplification using primer pairs that include TruSeq Illumina adapters and a 5′ multiplexing tag (TruSeq i7 index region indicated in bold). This amplification utilized a universal forward primer (5′ AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCGCACGTCTGCCGCGCTG 3′) and a unique reverse primer (5′ CAAGCAGAAGACGGCATACGAGAT **NNNNNN**GTGACTGGACTTCAGACGTGTGCTCTTCCGATCCAGGTTCTTGCGAACCTCAT 3′).

A single-step PCR amplification of sgID-BC regions was employed, proving to be a highly reproducible and quantitative method for determining the neoplastic cell count in each tumor. For each mouse, eight 100 μL PCR reactions per sample (4 μg DNA per reaction, 32 μg per mouse) were conducted using Q5 High-Fidelity 2 × Master Mix (New England Biolabs, M0494X). The resulting PCR products were purified with Agencourt AMPure XP beads (Beckman Coulter, A63881) using a double size selection protocol.

The concentration and quality of the purified libraries were assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, G2939BA). The libraries were pooled based on lung weight for even sequencing depth distribution, cleaned up, size-selected using AMPure XP beads, and sequenced on the Illumina® HiSeq 2500 platform, generating Paired-End 150 bp reads (PE150) (Novoegene). To enhance sequencing diversity and improve quality, 15%—25% PhiX control DNA was added to the library at the beginning of the sequencing reads.

## Quantification of tumor cell number from tumor barcode sequencing

High-quality PE150 reads were each aligned to the known reference sequence of the lentiviral construct encapsulating 10 nucleotides flanking the dual barcodes on either side. Alignment to the reference sequence was performed on both the forward and reverse reads using a Striped Smith-Waterman algorithm to ensure lossless extraction of putative barcodes from every read.[21] A prespecified scoring matrix including i) nucleotide match = +4, ii) nucleotide mismatch = −2, iii) gap open = −6, and iv) gap extension = −1 was used. Barcodes were then extracted from each of the forward and reverse read alignments with mismatching bases annotated as "N". Barcodes were then "dereplicated" (tallied), and barcode tallies were removed if the forward and reverse barcode lengths did not match, or the mean alignment score of the entire pileup was below 50%. This lossless dereplication of barcodes followed by filtering based on the global properties of the barcode pileup minimized biases that sequencing fidelity imparts on pileup quantity (*e.g.*, barcodes with homopolymer runs generally exhibit lower PHRED scores), while still ensuring that only true lentiviral barcodes are tallied.

The rate of recurrent read errors was estimated and used to remove spurious barcodes from pileups using an eight-parameter error model: six symmetric substitution rates to/from A, T, C, G, and N; and an Insertion and Deletion rate. These parameters were estimated once from the median rate of observed recurrent read errors descending from the 100 largest barcode pileups in every mouse in this study. The probability that every barcode in this study was a recurrent read error was then estimated by calculating a rate statistic λ from each barcode's nearest neighbor barcode size multiplied by the estimated error rate for the difference between the two barcodes in question (*e.g.*, substitution or InDel). A Poisson distribution was then used to estimate the likelihood that a pileup was a recurrent read error based on this rate statistic and pileups with a likelihood of over $10^{-10}$ were filtered from this study (Fig. S1B—S1D). This filtering probability has been used previously.[16,22]

Guanine-cytosine (GC) content bias during PCR amplification and the likelihood of "barcode collisions" were then estimated for every mouse and Lenti-sgRNA/mBC pool; however, minimal variation across the experiment was observed. A "barcode collision" rate constitutes the likelihood that two lentiviruses with the same barcode pair transduce different cells in the same mouse. The probability of barcode collisions can be estimated based on the frequency with which the same barcodes appear in separate mice in the study (high diversity barcodes appear only once in the entire study, whereas lower diversity barcodes may be observed in two or three mice). The probability of barcode collisions can bias growth estimates if Lenti-sgRNA pools are barcoded with varying efficiency; however, in this study, we did not observe appreciable differences in barcode diversity across sgRNAs. PCR amplification bias was also not observed, as seen previously.[16]

Absolute tumor numbers were then estimated for every non-filtered barcode pileup in every mouse by simply dividing the read number by the median read number observed for the three spike-in barcodes and multiplying by the known cell number of these barcodes ($1 \times 10^5$). A minimum cell cutoff of 400 cells yielded tumor size profiles that were highly reproducible across the technical replicates in our study, and minimized variation in median tumor size between the two inert sgRNAs (Neo1 and Neo3).

## Comprehensive statistical assessment of tumor progression using TuBa-seq

Previous TuBa-seq analyses found that a single summary statistic could not explain differences in tumor progression observed between different tumor genotypes. In particular, both the quantity of transduced lineages and the size spectrum of distinct genotypes reproducibly vary (*e.g.*, Stk11 loss dramatically increases mean tumor size without appreciably altering tumor number, while phosphatase and tensin (Pten) loss dramatically increases tumor number with only a moderate increase in mean tumor size).[23] Size spectrums then reproducibly differ in both the central tendency (mean/median) of growth and the preponderance of exceptionally large tumors, which cannot be explained by any single random process. Because malignancies are only observed at later time points (and most transduced lineages remain premalignant), it is imperative to quantify both the immediate effects of gene loss (mean and/or 50th − 95th percentile) and the ability of a genotype to potentiate clinically-relevant disease ("heavy-tail" tumor burden and/or 99th − 99.99th percentile).

Total tumor quantity, percentiles of tumor size from the 10th to 99.99th percentiles, and parametric summary statistics of size profiles were all calculated for every sgRNA pool in every mouse as described previously.[16] Central tendency was estimated using the maximum likelihood estimator (MLE) of the mean for a log-normal distribution, while deviating from a log-normal distribution in the large "heavy" tail of tumor sizes.[24] The minimum size cutoff for the Pareto tail of tumor burden was calculated relative to the median tumor size of sg*Inerts* and inferred to be 44.73 times the median for this study. All statistics for active sgRNAs were then divided by their respective statistic for sg*Inerts* within each mouse, and then averaged across all mice in a particular experimental cohort. *p*-values and confidence intervals for percentiles and parametric statistics were then calculated using two million iterations of the Bootstrap resampling method. *p*-values for tumor number (tumor initiation) were calculated using Fisher's exact test.

## Total RNA isolation protocol from mouse lung tissues and RNA sequencing

Tumor-bearing lung lobes from mice intubated with Lenti-sg*Egr1*/Cre or Lenti-sg*Inert*/Cre pools were mixed with TRIzol reagent (Invitrogen Life Technologies) and homogenized using a gentleMACS dissociator (Miltenyi Biotec). Subsequently, 1 mL of the lysate was treated with 0.2 mL of chloroform per 1 mL of TRIzol and vigorously shaken for 15−30 s, followed by incubation at room temperature for 2−3 min to induce phase separation. The mixture was then centrifuged at 12,000−16,000 *g* at 4 °C for 15 min. The RNA-containing aqueous phase was carefully transferred to a new tube, and RNA purification was performed using the Qiagen RNeasy Mini Kit (Qiagen #74104) according to the manufacturer's protocol. RNA quality was assessed based on RNA integrity numbers using a 2100 Bioanalyzer. Sequencing was performed using NovaSeq X Plus Series (PE150) reading 12 G raw data per sample.

## Histology and immunohistochemistry

Lung tissue samples were obtained from 1 to 3 mice per cohort, transduced with the Lenti-sgTSPool/Cre pool or individual sgRNAs (sg*Egr1* and sg*Inerts* pools). The collected lobes were fixed in 4% paraformaldehyde for 12 h, followed by storage in 1 × phosphate-buffered saline at 4 °C. Subsequently, tissues were paraffin-embedded and sectioned at the University Hospital Histology Core Facility into 4-μm sections. Hematoxylin and eosin staining was conducted for visualization, and tumor sizes were assessed by measuring the longest diameter of each tumor in the hematoxylin-eosin-stained sections using ImageJ software. Immunohistochemical staining was performed using specific antibodies: anti-TTF-1/Nkx2−1 (1:200, abcam, AB76013), anti-CD3 (1:750, AB135372), anti-CD4 (1:1000, AB183685), antiCD8 (1:1000, ab217344), anti-CD19 (1:1000, ab245235), anti-F4/80 (1:100, Invitrogen# 14-4801-85), and anti-Egr1 (1:200, ab6054). Statistical analysis was performed using ImageJ software, with *p*-values were calculated using the Mann−Whitney *U* test.

## RNA sequencing analysis

Raw sequence reads were processed to remove contaminant DNA, PCR duplicates, and adaptor sequences. Reads with a quality score below Q20 were excluded using CLC Genomics Workbench (v22.0.2). Paired-end reads were then assembled and aligned to the latest human genome reference (Homo sapiens, GRCh38). Transcript assembly and abundance quantification were conducted via StringTie.[25] After FPKM (fragments per kilobase of transcript per million mapped reads) normalization with Ballgown,[26] different gene expression analysis was conducted with edgeR's exact test (edgeR v4.2.1).[27] Pairwise comparisons were performed with edgeR's decideTestsDGE function, and a Benjamini-Hochberg-adjusted *p*-value cutoff of 0.05 was used to obtain differentially expressed genes (absolute $\log_2$ fold-change $\geq 1$). Gene Set Enrichment Analysis (GSEA) was performed using the murine hallmark gene set from MSigDB (v7.5.1)[28]; significantly enriched pathways had a

Benjamini-Hochberg-adjusted $p$-value $\leq 0.05$ and a minimum gene set size of 10. Data visualizations were conducted through R (v4.4.1) packages. Significantly up-regulated genes underwent Gene Ontology (GO) enrichment analysis using the Metascape tool. Statistically enriched GO terms were identified using cumulative hypergeometric $p$-values and enrichment scores. Then, a network-based approach was used to identify the most central genes in immunity-related pathways. Specifically, differentially up-regulated genes from the top 18 significantly enriched and immunity-related pathway clusters were analyzed for Maximal Clique Centrality (MCC) using the CytoHubba plug-in of Cytoscape. A comprehensive set of T-cell exhaustion markers was curated from multiple resources, including CellMarker 2.0 and 10X Genomics. Differential expression of 55 curated markers was assessed between Egr1-KO and control samples, and their expression profiles were visualized in a heatmap.

### *Egr1* knockout and expression analysis using publicly available DepMap and PCAWG datasets

The impact of EGR1 knockout was analyzed using RStudio with data from the DepMap CRISPRGeneEffect dataset, which was integrated with the model dataset to classify cancer types, and with the OmicsSomaticMutations dataset to identify cell lines harboring KRAS mutations. To determine the significance of EGR1 knockout effects, we applied cutoffs of $-0.5$ and 0.5, as suggested by DepMap. *EGR1* expression in clinical samples was evaluated using the Pan-Cancer Analysis of Whole Genomes (PCAWG) dataset. Expression levels in normal lung tissue were compared with those in LUAD by filtering samples based on histology and using DCC project codes to identify normal tissue. Statistical significance between the two groups was calculated using the Wilcoxon signed-rank test.

## Results

### Development of the first *in vivo* lung cancer model of TBX2 subfamily signaling-associated genes using multiplexed CRISPRko and lineage-tracing

To enable CRISPR/Cas9-mediated somatic genome editing within the context of an oncogenic *Kras* variant, we crossed mice with Cre/lox-activated alleles of $KRAS^{G12D}$ ($Kras^{LSL-G12D/+}$) and Cre/lox-activated Cas9 ($R26^{LSL-Cas9-eGFP/+}$), hereafter named a KC mouse model.[16,29] Tumors were initiated through intratracheal intubation with barcoded pooled-Lenti-sgRNA/Cre vectors targeting the four TBX2 genes (*Tbx2*, *Tbx3*, *Tbx4*, and *Tbx5*) and four of their associated effector genes identified previously through our *in vitro* transcriptome profiling (*Egr1*, *Chd2*, *Tnfaip3a*, and *Atf3*) (Fig. 1A and B).[11] The utilized lenti-sgRNA-Pool/Cre pool included control genome-targeting inerts (sg*Neo1* and sg*Neo2*), and positive and negative controls targeting the known tumor suppressor gene *Rb1* and the essential gene *Pcna*, respectively. Each gene was targeted with two unique Lenti-sgRNA/Cre vectors optimized for on-target cutting specificity within the first three exons of the targeted gene (Table S1). To investigate the role of these genes at both early and late tumorigenesis stages, cohorts of mice were evaluated at 6 weeks and 20 weeks after intubation, respectively (Fig. 1B). Lung weights and histological examination revealed fewer visible tumors at 6 weeks, despite these mice receiving higher viral titers (Fig. 1C and D; Fig. S1A). Lung tumors observed at both time points included hyperplastic lesions and adenomas with foci of high-grade/malignant features, all of which stained positively for the lung lineage-defining transcription factor NKX2.1/TTF-1 (Fig. 1C).

### Multiplexed quantification of TBX2 gene function in *Kras*-driven lung tumors

To precisely quantify the growth effects of each gene knockout within the same mouse at all potential tumor sizes (from tens of cells to millions), we stably labeled the genetically diverse tumors with lentiviral-mediated DNA barcodes (termed Tumor Barcoding, TuBa-seq). Tumor sizes and quantities were determined *en masse* by analyzing genomic DNA extracted from bulk tumor-bearing lung tissue, followed by deep sequencing of the double barcode region, which identifies both the short guide RNA (sgID) and descendants ("million" BarCode, mBC) of every transduced cell. Sequencing performed to an average depth of $>10^7$ reads per mouse allowed us to reliably quantify tumor sizes below 400 cells. To translate barcode read tallies to absolute tumor sizes, we spike-in DNA barcodes of known cell number before cell lysis. After barcodes are tallied, we then eliminate spurious tumors and potential technical replicates by generating a statistical model of sequencing errors and barcode diversity (Fig. S1B–S1D). We assessed the impact of knocking out each of the potential tumor suppressor genes by analyzing the distribution of absolute cell numbers for each gene knockout (Fig. 2A). Consistent with histological evaluations, the mean total cell number initiated by either Lenti-sg*Inerts* or Lenti-sgTSGs was significantly higher in the 20-week cohort compared with the 6-week cohort (Fig. S1E). The average total cell number produced by a single sg*Inert* was significantly smaller than the combined mean cell number produced by our studied knockouts, indicating an overall marked acceleration in cellular proliferation.

Lineage tracing uncovered a wide spectrum of effects on tumorigenesis and variability in tumor sizes with a unique profile of growth effects for each tumor suppressor gene knockout (Fig. 2A). Both the mean estimate and 95th percentile of tumor sizes of each gene knockout was highly correlated between the two sgRNAs targeting a gene ($r = 0.87$ for mean estimate; Fig. 2B; Fig. S1F). Strong correlation of the percentile spectrums of the two Inert sgRNAs for replicates at both 6 and 20 weeks demonstrates excellent reproducibility of our growth profiles (Fig. 2C). Similarly, A quantile–quantile (Q–Q) comparison of each gene knockout against the wild-type (*Kras*-only) tumor size distribution in each mouse summarizes growth effects across all tumor sizes (Fig. 2D; Fig. S2A). As cataloged previously via TuBa-seq,[17] a gene's effects on tumorigenesis reproducibly alter the number of tumor lineages transduced by the same lentiviral pool, the mean/median size of

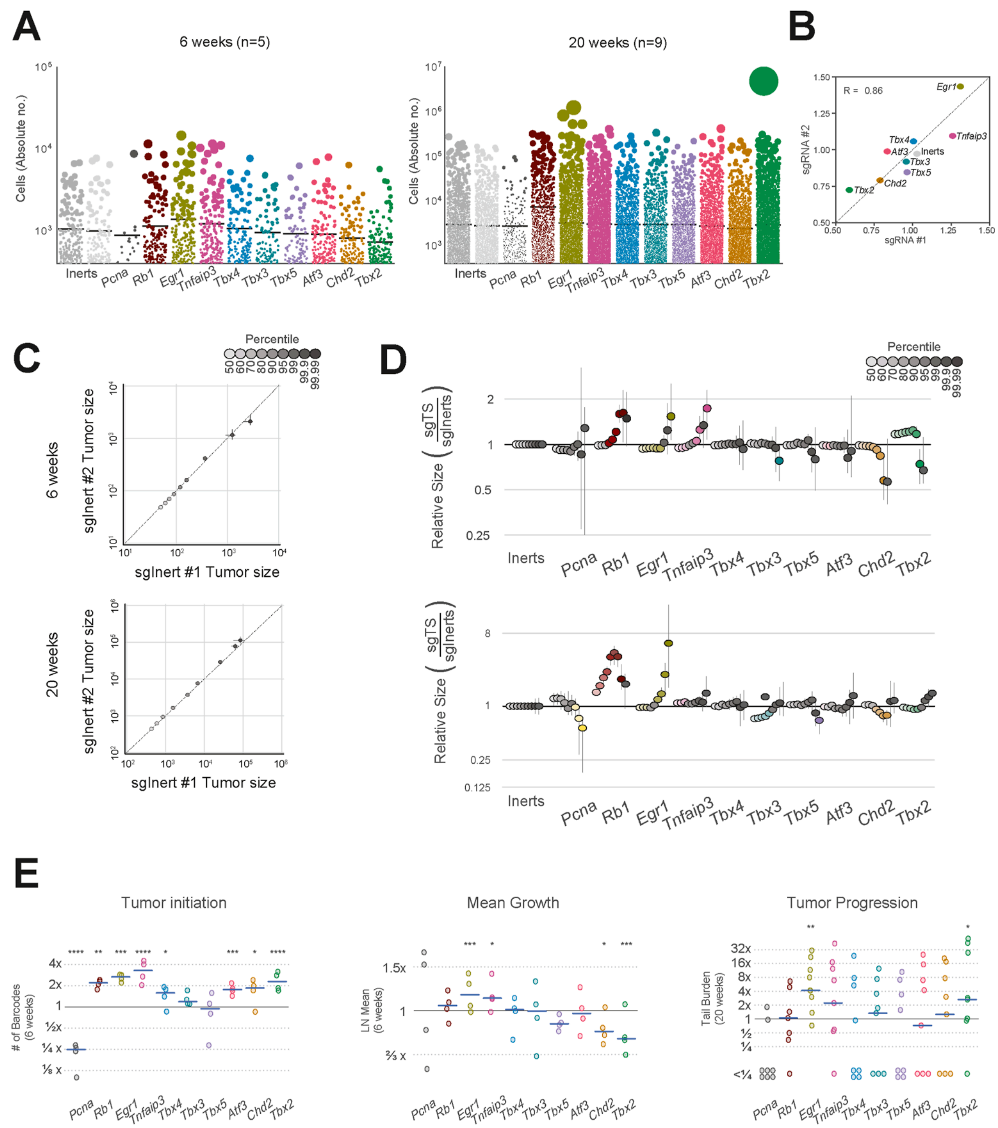

**Figure 2** Multiplexed somatic CRISPR–Cas9 genome editing reveals that loss of *Egr1* has the most pronounced and consistent effect on tumor initiation, growth, and progression in *Kras*-driven lung cancer. **(A)** Individual tumor sizes ordered by mean effect of targeted gene knockout in KC mice at 6 weeks and 20 weeks after tumor initiation. The number of mice in each cohort is as indicated. Each dot represents a tumor. The area of each dot is proportional to the number of cancer cells in each tumor. **(B)** Mean effect of gene knockout of each sgRNA in KC mice at 6 weeks after tumor initiation. LN mean is a maximum likelihood estimator (MLE) of the mean tumor size under a log-normal sampling distribution. Each gene in this study was targeted by two high-specificity sgRNAs, imparting consistent growth effects (Pearson's $r = 0.86$, $p = 0.0029$; see *Materials and Methods*). **(C)** Quantile–Quantile (Q–Q) plot of tumor size distributions of the top 50th – 99.99th percentiles of the two Inert sgRNAs at both 6 and 20 weeks demonstrates reproducibility of size profiles. **(D)** Analysis of relative tumor sizes in *KRAS*-driven tumors at 6 and 20 weeks after tumor initiation. Relative size of tumors for each targeted gene (sgTS) is divided by the respective sg*Inerts* percentile and merged across replicate mice. Percentiles significantly different from sg*Inerts* are in color ($p < 0.05$, two-sided bootstrap resampling). The darker the shade of color, the larger the percentile, as shown in the legend in gray scale. Error bars denote 95% confidence intervals also determined by bootstrap sampling. **(E)** Summary of three distinct growth phenotype profiles. For each genotype, the number of tumors observed at six weeks after tumor initiation, alongside MLE of mean size at six weeks (tumor growth; see *Materials and Methods*), and the size of the largest tumors at 20 weeks (advanced progression) is depicted. All statistics are divided by their respective values for sg*Inerts*. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$ (two-sided bootstrap resampling).

a transduced tumor, and the probability of exceptionally large tumors. These exceptionally large tumors cannot be explained by a single-step Markov process, nor mode of gene-editing (Floxed-allele or CRISPRko), and are consistent with (epi)genetic progression over the time course.[16] Hence, we quantified tumor progression via three distinct, previously-vetted progression measures: i) tumor initiation effect (number of barcodes observed), ii) mean growth effect (a MLE of the mean based on a log-normal distribution, LN mean; Fig. S2B), and (iii) effect on advanced progression (total burden of tumors larger than sizes expected from a log-normal distribution; see *Materials and Methods*). Summary statistics for each gene were reported relative to wild-type summary statistics within each mouse, thereby controlling for extensive variability in tumor growth observed between mice (Fig. 2E).[16]

Tumor necrosis factor alpha-induced protein 3 (*Tnfaip3*)-deficient cells showed a marked increase in tumor size after 6 weeks of growth, suggesting a role in early tumor growth but not necessarily in advanced progression ($p < 0.05$ for 95th − 99th percentiles, and LN mean, bootstrap resampling; see *Materials and Methods*) (Fig. 2D). Tnfaip3, also known as A20, is a key anti-inflammatory enzyme and a critical regulator of inflammation homeostasis. Prior studies have demonstrated that the intrinsic loss of A20 in tumor cells markedly enhances lung tumorigenesis and impairs $CD8^+$ T cell-mediated immune surveillance in both patients and murine models.[30]

The profile of *Tbx2*-deficient cells exhibited a unique pattern, with an increase in tumor initiation capacity and relative sizes of smaller percentiles, accompanied by a significant decrease in overall LN mean at 6 weeks and the largest percentiles (Fig. 2D). Furthermore, this knockout led to the formation of some of the largest tumors observed in our screen at 20 weeks. Countervailing context-dependent effects of gene knockouts on growth have been observed previously and are consistent with known *Tbx2* biology. Inappropriate activation of *Tbx2* is thought to contribute to tumor progression by overriding senescence, thereby sustaining tumor growth. *TBX2* overexpression has been observed in pancreatic, colorectal, melanoma, endometrial, ovarian, and cervical cancers, where it contributes to tumorigenesis.[15] Conversely, other lung and skin cancer models have shown that *Tbx2* overexpression inhibits cell growth while being associated with increased resistance of tumor cells to the anti-cancer drug cisplatin, consistent with a highly contextual role of *Tbx2* in cancer progression.[31]

Loss of *Tbx3*, *Tbx4*, *Tbx5*, and activating transcription factor 3 (*Atf3*) exhibited only marginal growth effects in *Ras*-driven tumors, similar to their marginal effects on absolute cell number (Fig. 2D and E). Chromodomain helicase DNA binding protein 2 (*Chd2*) loss consistently resulted in suppressed cellular proliferation with a reduction in total cell number at top percentiles and a significant decrease in LN mean at the studied time point. In general, inactivation of CHD family members has been implicated in various human cancers. Notably, CHD2 has been suggested to play a role in preventing breast cancer initiation. In lung cancer, *CHD2* mRNA expression has been linked to cancer stage, particularly in lung squamous cell carcinoma.[32]

Lastly, deactivation of *Egr1*, a *Tbx2* partner and a direct regulator of key tumor suppressors such as *Tgfβ1*, *Pten*, and *Tp53*, increased mean growth rates at both early and late time points and resulted in exceptionally large tumors at 20 weeks after initiation ($\sim 5 \times$ size increase, two sample $t$-test, $p < 0.05$), surpassing even the effect of *Rb1* knockout.[33] Strikingly, only *Egr1* loss significantly promoted all three growth phenotypes, exhibiting this screen's strongest advanced progression phenotype (Fig. 2D and E; Fig. S2B).

## Immune pathway modulation and tumor progression following *Egr1* suppression in *Kras*-driven lung cancer

To validate the tumor-suppressive role of *Egr1* and investigate its mechanism of action in a *Kras*-driven background, we generated *Egr1*-deficient tumors in KC mice by delivering Lenti-sg*Egr1*-mBC/Cre pools and compared them with *Egr1*-intact control tumors induced in KC mice using Lenti-sgInert-mBC/Cre pools (Fig. 3A). Twenty weeks after tumor initiation, *Egr1*-deficient mice developed larger tumors compared with *Egr1* controls (Fig. 3B; Fig. S3A).

To investigate the molecular mechanisms of *Egr1*-driven tumor growth, whole transcriptome comparison between tumor-bearing *Egr1*-knockout lungs and controls was performed (Fig. S3B). 421 genes were significantly deregulated in *Egr1*-knockout lung tissues compared with control samples. Unsupervised hierarchical clustering of gene expression profiles revealed distinct differential expression patterns between the groups (Fig. 3C). Nearly all the top up-regulated genes were closely linked to immune responses, inflammation, and chemokine signaling, including *Il17a*, *Itln1*, *Cd177*, *Gp2*, *Cxcl5*, and *Cxcl13*. Gene Ontology (GO) pathway analysis of the differentially up-regulated genes corroborated the immune-modulatory effects of *Egr1* loss, with leukocyte migration and activation emerging as the most enriched ontology clusters (Fig. 3D). Building on these findings, MCC analysis further pinpointed the top 12 hub genes within the most enriched immune-related pathways. These genes played critical roles in immune regulation and pathway interconnectivity across the eight primary immune ontology categories (Fig. 4A). Furthermore, GSEA revealed significant enrichment in three immune and inflammatory signaling hallmarks: allograft rejection, interferon gamma response, and interferon alpha response, collectively suggesting that *Egr1* deficiency leads to a more immunogenic tumor phenotype. Conversely, although less pronounced, several hallmark tumorigenesis pathways, including Myc, mTOR, and reactive oxygen species (ROS) signaling, were down-regulated in these tumors (Fig. S3C).

We employed three complementary approaches to further explore the impact of *Egr1* knockout on lung tumorigenesis and the role of the tumor microenvironment in this process. First, using the PCAWG dataset, we demonstrated significant suppression of *EGR1* expression in clinical adenocarcinoma tissues compared with normal lung tissues (Fig. S3D). Next, we quantified the growth effects of *EGR1* knockout in human *KRAS*-driven LUAD cell lines and observed no significant growth differences in the absence of an immune microenvironment (Fig. 4B; Table S2). Additionally, immunohistochemical staining of lung tissues revealed a marked increase in T cell infiltration in *Egr1*-KO tumor tissues, with robust infiltration of both helper T cells and cytotoxic T cells, indicating an immune response linked to *Egr1* loss (Fig. 4C). In contrast, macrophage staining intensity was similar between control and knockout samples, and B cell staining, marked by CD19, was comparable in both groups, although more aggregates were observed in the control samples (Fig. 4C).

To further explore the molecular signature of T cell status in the tumor microenvironment of *Egr1*-deficient

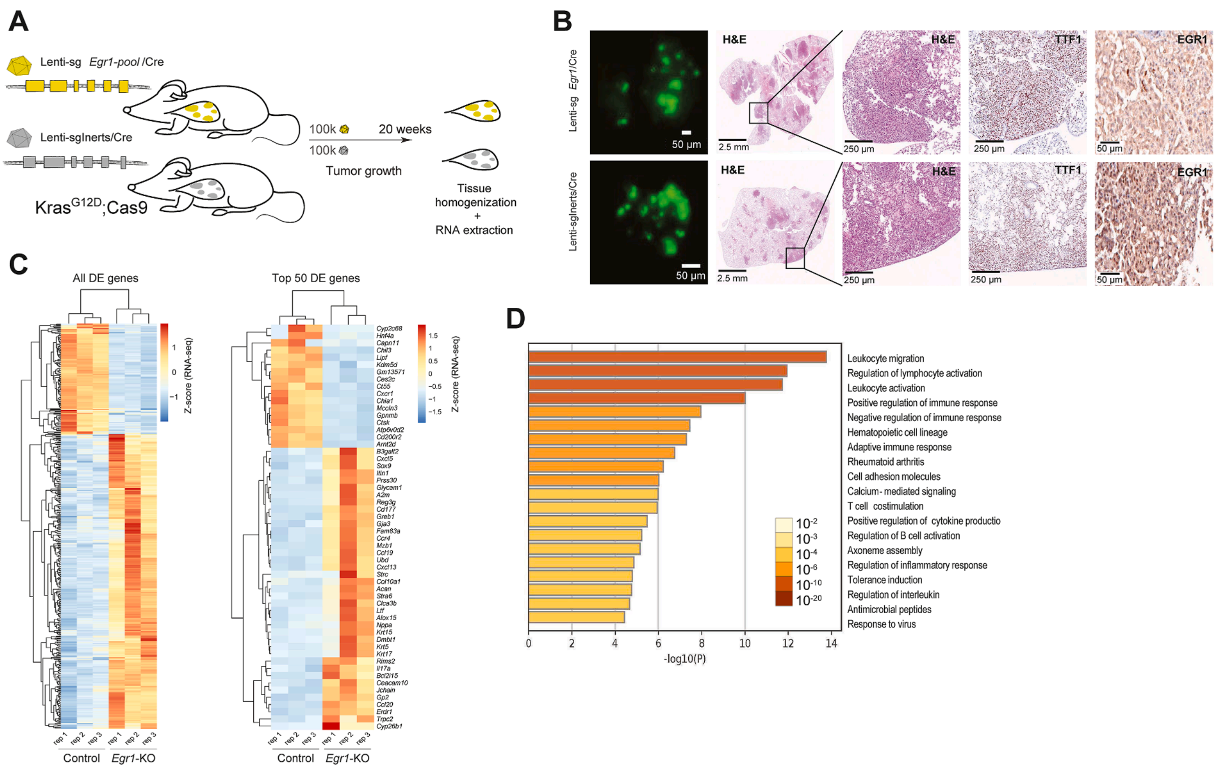


**Figure 3** Loss of *Egr1* promotes tumor progression, immune modulation, and differential gene expression in *Kras*-driven lung cancer. **(A)** Schematic illustration of the experimental design for *Egr1* knockout versus control. KC mice were administered a pool of two sgRNAs targeting *Egr1* ($n = 3$) or a pool of control sg*Inerts* ($n = 3$). **(B)** Representative images of lung lobes from mice at 20 weeks after tumor initiation via Lenti-sg*Egr1*-pool/Cre or Lenti-sg*Inerts*. Images include fluorescence views under a dissecting microscope, hematoxylin-eosin-stained sections, TTF1-stained sections to confirm lung adenomatous origin, and Egr1 staining to confirm knockout efficiency. **(C)** Bulk RNA-sequencing analysis was conducted on lung tissues at 20 weeks after tumor initiation in *Egr1* knockout and control mice. The left panel displays a heatmap of 421 differentially expressed genes (DEGs) identified with a false discovery rate-adjusted *p*-value <0.05, illustrating two distinct expression clusters. The right panel highlights a heatmap of the top 50 most deregulated genes, showing clear differences between the two groups. **(D)** Gene Ontology analysis was performed using Metascape on significantly up-regulated genes between *Egr1* knockout and control mice. Statistically enriched terms were identified based on cumulative hypergeometric *p*-values and enrichment scores. The top 20 most significant terms, predominantly associated with immune processes, are visualized.

samples, we focused on gene sets associated with T cell exhaustion. *Egr1* knockout samples exhibited substantial differential expression, including up-regulation of key T cell exhaustion markers, such as *Pdcd1* (PD-1), *Cd244* (2B4), *Lag3*, and *Ctla4* (Fig. 4E). These findings suggest a potential transcriptional shift suggestive of T cell exhaustion in the absence of *Egr1*. However, it is important to note that further detailed analysis using single-cell RNA sequencing or spatial transcriptomics would provide a more comprehensive understanding of the T cell status within the tumor microenvironment, which we plan to address in future studies.

## Discussion

Most putative driver mutations are altered in <10% of tumors, highlighting the need to understand the functional importance of additional actionable drivers that may be infrequently altered at the genetic level, but often deregulated through expression changes.[29] Furthermore, the impact of co-occurring tumor suppressor gene alterations is often overlooked in the molecular classification of tumors. Our study addresses the limitations of current methodologies, which predominantly focus on single-driver oncogenic mutations, thereby neglecting the broader landscape of genetic and epigenetic changes in putative tumor suppressors that could contribute to tumorigenesis. In this context, we applied for the first time TuBa-seq technology to systematically investigate the role of *Tbx2* genes and their effectors *in vivo*. Contrary to gene-centric models of tumor progression, we found that epigenetically dysregulated genes, such as the Tbx2 subfamily, can causally drive or slow tumor progression when deleted. This model provided highly precise and detailed quantification of tumor growth following the inactivation of our targeted genes in *Kras*-driven lung tumors.

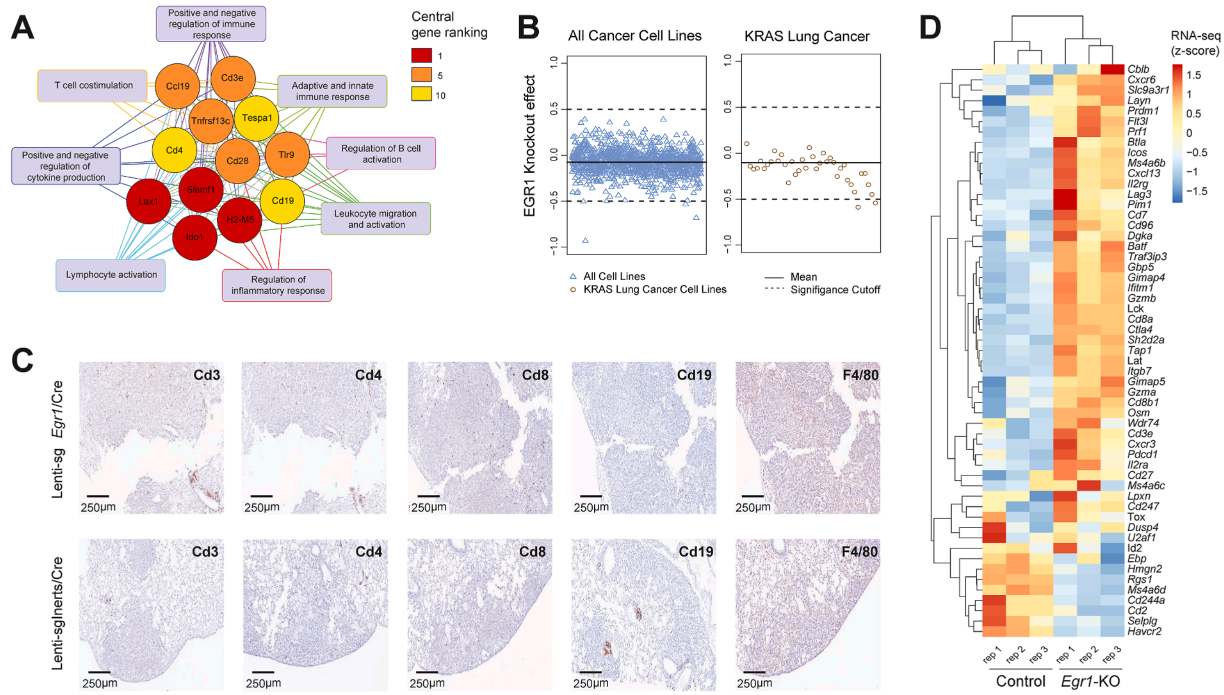


**Figure 4** Impact of *Egr1* knockout on *in vitro* lung cancer proliferation, *in vivo* T cell infiltration, and T cell exhaustion in the tumor microenvironment. **(A)** The immune network and top 12 hub genes are visualized. From the top 30 most significantly up-regulated pathway clusters, 18 immunity-related terms and their associated genes were selected for Maximal Clique Centrality analysis using the CytoHubba plugin in Cytoscape. Gene interactions within the network indicate their participation in shared pathways. The background color of each gene box represents its centrality ranking: four genes co-ranked 1st (red), five genes co-ranked 5th (orange), and three genes co-ranked 10th (yellow). The 16 pathways associated with these 12 hub genes were further consolidated into 8 primary categories, represented by blue rectangular boxes. **(B)** Analysis of the DepMap CRISPR—Cas9 cancer dataset, evaluating the effect of *Egr1* knockout in 342 human cancer cell lines, including independent assessments in *KRAS*-driven lung cancer lines, showed no significant impact on cell proliferation in either case. **(C)** Immunohistochemical staining of Cd4, Cd8, Cd3, Cd19, and F4/80 in *Egr1*-deficient lung tumors compared with control tumors. Scale bars are provided for each image. **(D)** The heatmap illustrates the expression patterns of 55 curated T-cell exhaustion markers (CellMarker 2.0 and 10X Genomics) in control and *Egr1*-deficient tumors. Differential expression analysis reveals that most exhaustion markers are up-regulated in *Egr1*-deficient samples. The color scale represents relative expression levels, ranging from red (high expression) to blue (low expression).

In this study, genes associated with the TBX2 subfamily exhibited distinct roles at different stages of tumor progression. *Chd2* knockout inhibited tumor growth, while *Tnfaip3* knockout promoted tumor initiation and overall growth. Interestingly, cells deficient in *Tbx3*, *Tbx4*, *Tbx5*, and *Atf3* demonstrated minimal effects on tumorigenesis within our experimental model. Consistent with most studies, *Tbx2*-deficient cells exhibited suppressed growth capacity, yet its knockout significantly contributed to both tumor initiation and advanced progression—two aspects of TBX2's role in tumorigenesis that remain understudied. Tbx2 is known to physically interact and suppress the transcriptional activity of *Egr1* (the most significant tumor suppressor gene identified in our study) (Table S3). *Egr1* is a zinc-finger transcription factor with multifunctional roles in proliferation, stress responses, and apoptosis.[34] In cancers, such as rhabdomyosarcoma and breast cancer, TBX2 exploits this interaction to target several carcinogenic genes, including N-myc downstream-regulated gene 1 (NDRG1), a protein involved in cell differentiation, apoptosis, and senescence (Fig. 1A).[23,24]

Beyond its relationship with TBX2, *EGR1* has been extensively studied in cancer biology. It is activated through the mitogen-activated protein kinase (MAPK) signaling pathway in response to various stimuli, such as growth factors, tumor necrosis factor, hypoxia, inflammatory signals, ionizing radiation, and ROS. Once activated, *Egr1* can either promote or suppress the expression of its target genes, influencing transcriptional regulation. As a tumor suppressor in gliomas and melanocytomas, Egr1 up-regulates p21Waf1/Cip1 to induce apoptosis.[35] Additionally, *Egr1* enhances tumor cell death by directly up-regulating tumor suppressors like NSAID-activated gene 1 (*NAG1*) and *PTEN*.[33] In the context of immune regulation, *Egr1* has been the subject of conflicting reports, with its role oscillating between immunostimulatory and immunosuppressive effects depending on the cellular and environmental context. In inflammatory lung diseases, EGR1 is recognized as a master regulator of transcription, promoting the expression of genes critical for immune signaling and inflammatory cell activation.[36,37] Conversely, within macrophages, EGR1 has been found to interact indirectly with the nucleosome remodeling and deacetylase (NuRD) complex, leading to chromatin deacetylation and the suppression of inflammatory enhancer activation, thereby curbing excessive inflammatory responses.[36] In the context of *KRAS*-driven cancer, few studies have explored the role of EGR1. Notably, EGR1 is known to be transcriptionally activated by ERK1/2, a key downstream effector of oncogenic KRAS signaling.[38] Supporting a tumor-suppressive role, a study using the LSL$^{KrasG12D/+}$; Pdx1$^{-Cre}$ mouse model of pancreatic cancer demonstrated that δ-tocotrienol treatment induced nuclear EGR1 expression and triggered apoptosis in early ductal lesions, highlighting a protective function for EGR1 in KRAS-mutant tumors.[39] Despite its importance, EGR1 remains largely understudied in lung cancer, with existing research primarily limited to clinical surveys and *in vitro* experiments.[40,41] For example, *EGR1* expression has been shown to predict *PTEN* levels and patient survival after surgical resection in non-small cell lung cancer, with lower *EGR1* levels correlating with poorer outcomes.[41] In this study, we present the first lung-specific *Egr1*-knockout mouse model, emphasizing its critical role at various stages of LUAD development, particularly in the context of *Kras*-driven tumorigenesis. *Egr1*-deficient cells demonstrated enhanced tumor initiation, growth, and progression, with effects exceeding those observed with the loss of *Rb1*, a key tumor suppressor in LUAD. Our transcriptomic analysis revealed that the tumorigenic effects of Egr1 deficiency were linked to an up-regulation of immune-related genes and a molecular signature suggestive of T cell exhaustion in the tumor microenvironment. However, further investigation is needed to fully understand how *Egr1* deficiency in epithelial cells contributes to the onset or exacerbation of T cell exhaustion, offering new insights into the complex immune dynamics within tumors.

In conclusion, our study identifies *Egr1* as a key player in lung cancer progression and immune modulation, positioning it as a promising target for immunotherapy or a potential biomarker. We note that this discovery would not have been made were this study limited to i) *in vitro* or ii) clinical exome-level sequencing analyses, or iii) a single gene within the TBX2 signaling cascade. Hence, this study underscores the indispensable value of systematic *in vivo* genetic models of cancer, which provide a more comprehensive understanding of gene function and mechanisms.

## CRediT authorship contribution statement

**Athar Khalil:** Conceptualization, Writing – original draft, Formal analysis, Data curation, Writing – review & editing. **Trang Dinh:** Formal analysis, Data curation. **Meaghan Parks:** Formal analysis. **Rebecca C. Obeng:** Formal analysis. **Berkley Gryder:** Formal analysis. **Adam Kresak:** Resources. **Yuxiang Wang:** Data curation. **Jeff Maltas:** Formal analysis. **Madeline Bedrock:** Methodology. **Xiangzhen Wei:** Methodology. **Zachary Faber:** Formal analysis. **Mira Rahm:** Methodology. **Jacob Scott:** Supervision, Writing – review & editing. **Thomas LaFramboise:** Writing – review & editing, Supervision. **Zhenghe Wang:** Funding acquisition, Conceptualization, Writing – review & editing, Supervision. **Christopher McFarland:** Funding acquisition, Formal analysis, Writing – review & editing, Conceptualization, Methodology.

## Conflict of interests

Zhenghe Wang is the member of *Genes & Diseases* Editorial Board. To minimize bias, he was excluded from all editorial decision-making related to the acceptance of this article for publication. The remaining authors declare no conflict of interests.

## Funding

This work was supported by grants from the US National Institutes of Health (No. R01CA271540, R00CA226506 to C.M.; R01CA196643, R01CA264320, R01CA260629, P50CA150964, P30 CA043703 to Z.W.).

## Acknowledgements

The authors gratefully acknowledge the Light Microscopy Imaging Core and the Histology and Tissue Procurement Facility at Case Western Reserve University for their support.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gendis.2025.101840.

## References

1. Leiter A, Veluswamy RR, Wisnivesky JP. The global burden of lung cancer: current status and future trends. *Nat Rev Clin Oncol*. 2023;20(9):624–639.
2. Papaioannou VE. The T-box gene family: emerging roles in development, stem cells and cancer. *Development*. 2014;141(20):3819–3833.
3. Wilson V, Conlon FL. The T-box family. *Genome Biol*. 2002;3(6). REVIEWS3008.
4. Gibson-Brown JJ, Agulnik IS, Silver LM, Papaioannou VE. Expression of T-box genes Tbx2-Tbx5 during chick organogenesis. *Mech Dev*. 1998;74(1–2):165–169.
5. Lüdtke TH, Farin HF, Rudat C, et al. Tbx2 controls lung growth by direct repression of the cell cycle inhibitor genes Cdkn1a and Cdkn1b. *PLoS Genet*. 2013;9(1):e1003189.
6. Lüdtke TH, Rudat C, Wojahn I, et al. Tbx2 and Tbx3 act downstream of Shh to maintain canonical Wnt signaling during branching morphogenesis of the murine lung. *Dev Cell*. 2016;39(2):239–253.
7. Arora R, Metzger RJ, Papaioannou VE. Multiple roles and interactions of Tbx4 and Tbx5 in development of the respiratory system. *PLoS Genet*. 2012;8(8):e1002866.
8. Khalil AA, Sivakumar S, Lucas FAS, et al. *TBX2* subfamily suppression in lung cancer pathogenesis: a high-potential marker for early detection. *Oncotarget*. 2017;8(40):68230–68241.
9. Nehme E, Rahal Z, Sinjab A, et al. Epigenetic suppression of the T-box subfamily 2 (*TBX2*) in human non-small cell lung cancer. *Int J Mol Sci*. 2019;20(5):1159.
10. Khalil A, Dekmak B, Boulos F, et al. Transcriptomic alterations in lung adenocarcinoma unveil new mechanisms targeted by the TBX2 subfamily of tumor suppressor genes. *Front Oncol*. 2018;8:482.
11. Khalil AA, Nemer G. *The TBX2 Subfamily of Transcription Factors: An Emerging Role in Lung Carcinogenesis*. New York, USA: American University of Beiru; 2018.
12. Li S, Luo X, Sun M, et al. Context-dependent T-BOX transcription factor family: from biology to targeted therapy. *Cell Commun Signal*. 2024;22(1):350.

13. Niu G, Hao J, Sheng S, Wen F. Role of T-box genes in cancer, epithelial-mesenchymal transition, and cancer stem cells. *J Cell Biochem*. 2022;123(2):215–230.
14. Yu J, Ma X, Cheung KF, et al. Epigenetic inactivation of T-box transcription factor 5, a novel tumor suppressor gene, is associated with colon cancer. *Oncogene*. 2010;29(49): 6464–6474.
15. Wansleben S, Peres J, Hare S, Goding CR, Prince S. T-box transcription factors in cancer biology. *Biochim Biophys Acta*. 2014;1846(2):380–391.
16. Rogers ZN, McFarland CD, Winters IP, et al. A quantitative and multiplexed approach to uncover the fitness landscape of tumor suppression *in vivo*. *Nat Methods*. 2017;14(7):737–742.
17. Jackson EL, Willis N, Mercer K, et al. Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev*. 2001;15(24):3243–3248.
18. Platt RJ, Chen S, Zhou Y, et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell*. 2014;159(2): 440–455.
19. Vandivort TC, An D, Parks WC. An improved method for rapid intubation of the *Trachea* in mice. *JoVE J*. 2016;(108):53771.
20. Hill AJ, McFaline-Figueroa JL, Starita LM, et al. On the design of CRISPR-based single-cell molecular screens. *Nat Methods*. 2018;15(4):271–274.
21. Zhao M, Lee WP, Garrison EP, Marth GT. SSW library: an SIMD Smith-Waterman C/C++ library for use in genomic applications. *PLoS One*. 2013;8(12):e82138.
22. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: high-resolution sample inference from illumina amplicon data. *Nat Methods*. 2016;13(7):581–583.
23. Cai H, Chew SK, Li C, et al. A functional taxonomy of tumor suppression in oncogenic KRAS-driven lung cancer. *Cancer Discov*. 2021;11(7):1754–1773.
24. Alstott J, Bullmore E, Plenz D. Powerlaw: a python package for analysis of heavy-tailed distributions. *PLoS One*. 2014;9(1): e85777.
25. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33 (3):290–295.
26. Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat Biotechnol*. 2015;33(3): 243–246.
27. Robinson MD, McCarthy DJ, Smyth GK. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–140.
28. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417–425.
29. Rogers ZN, McFarland CD, Winters IP, et al. Mapping the *in vivo* fitness landscape of lung adenocarcinoma tumor suppression in mice. *Nat Genet*. 2018;50(4):483–486.
30. Breitenecker K, Homolya M, Luca AC, et al. Down-regulation of A20 promotes immune escape of lung adenocarcinomas. *Sci Transl Med*. 2021;13(601):eabc3911.
31. Wang B, Lindley LE, Fernandez-Vega V, Rieger ME, Sims AH, Briegel KJ. The T box transcription factor TBX2 promotes epithelial-mesenchymal transition and invasion of normal and malignant breast epithelial cells. *PLoS One*. 2012;7(7): e41355.
32. Lv Y, Lin W. Comprehensive analysis of the expression, prognosis, and immune infiltrates for CHDs in human lung cancer. *Discov Oncol*. 2022;13(1):29.
33. Inoue K, Fry EA. Tumor suppression by the EGR1, DMP1, ARF, p53, and PTEN network. *Cancer Invest*. 2018;36(9–10): 520–536.
34. Redmond KL, Crawford NT, Farmer H, et al. T-box 2 represses NDRG1 through an EGR1-dependent mechanism to drive the proliferation of breast cancer cells. *Oncogene*. 2010;29(22): 3252–3262.
35. Wang B, Guo H, Yu H, Chen Y, Xu H, Zhao G. The role of the transcription factor EGR1 in cancer. *Front Oncol*. 2021;11: 642547.
36. Trizzino M, Zucco A, Deliard S, et al. EGR1 is a gatekeeper of inflammatory enhancers in human macrophages. *Sci Adv*. 2021; 7(3):eaaz8836.
37. Gururajan M, Simmons A, Dasu T, et al. Early growth response genes regulate B cell development, proliferation, and immune response. *J Immunol*. 2008;181(7):4590–4602.
38. Weatherdon L, Stuart K, Cassidy M, et al. Reporter cell lines to screen for inhibitors or regulators of the KRAS-RAF-MEK1/2-ERK1/2 pathway. *Biochem J*. 2024;481(6):405–422.
39. Wang C, Husain K, Zhang A, et al. EGR-1/Bax pathway plays a role in vitamin E $\delta$-tocotrienol-induced apoptosis in pancreatic cancer cells. *J Nutr Biochem*. 2015;26(8):797–807.
40. Levin WJ, Press MF, Gaynor RB, et al. Expression patterns of immediate early transcription factors in human non-small cell lung cancer. The Lung Cancer Study Group. *Oncogene*. 1995;11 (7):1261–1269.
41. Ferraro B, Bepler G, Sharma S, Cantor A, Haura EB. EGR1 predicts PTEN and survival in patients with non-small-cell lung cancer. *J Clin Oncol*. 2005;23(9):1921–1926.
42. Ku HC, Cheng CF. Master regulator activating transcription factor 3 (ATF3) in metabolic homeostasis and cancer. *Front Endocrinol*. 2020;11:556.
43. Wang Z, Yan C. Emerging roles of ATF3 in the suppression of prostate cancer. *Mol Cell Oncol*. 2015;3(1):e1010948.
44. Jarosz M, Olbert M, Wyszogrodzka G, Młyniec K, Librowski T. Antioxidant and anti-inflammatory effects of zinc. Zinc-dependent NF-$\kappa$B signaling. *Inflammopharmacology*. 2017;25 (1):11–24.
45. Khan SF, Damerell V, Omar R, et al. The roles and regulation of TBX3 in development and disease. *Gene*. 2020;726:144223.
46. Oh TJ, Adhikari A, Mohamad T, Althobaiti A, Davie J. TBX3 represses TBX2 under the control of the PRC2 complex in skeletal muscle and rhabdomyosarcoma. *Oncogenesis*. 2019;8 (4):27.
47. Li J, Ballim D, Rodriguez M, et al. The anti-proliferative function of the TGF-$\beta$1 signaling pathway involves the repression of the oncogenic TBX2 by its homologue TBX3. *J Biol Chem*. 2014; 289(51):35633–35643.
48. Chang F, Xing P, Song F, et al. The role of T-box genes in the tumorigenesis and progression of cancer. *Oncol Lett*. 2016;12 (6):4305–4311.