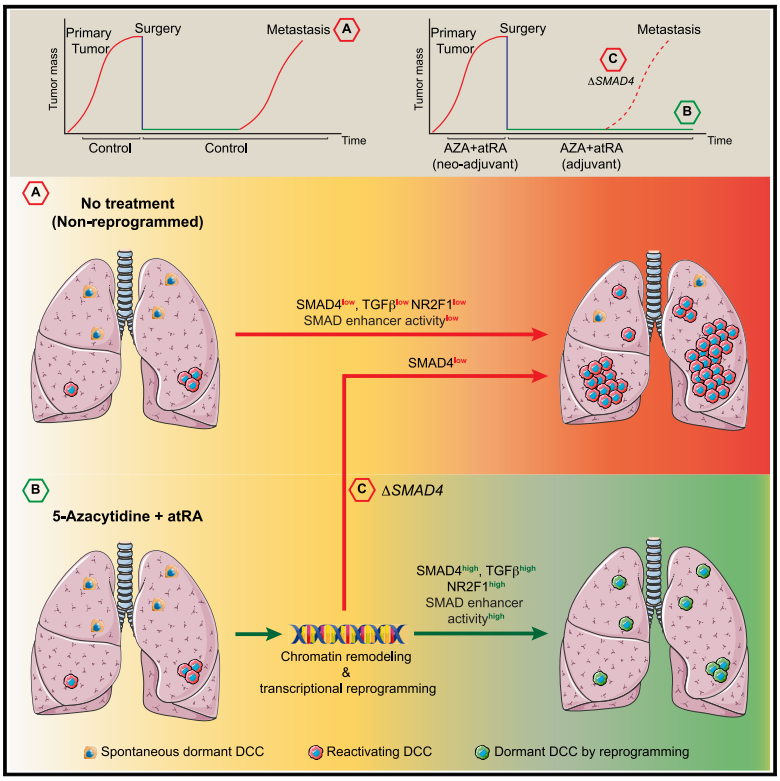
Cell Reports

Article

5-Azacytidine- and retinoic-acid-induced reprogramming of DCCs into dormancy suppresses metastasis via restored TGF-β-SMAD4 signaling

Graphical abstract



Highlights

- AZA+atRA reprogramming induces and maintains dormancy to suppress metastatic outgrowth
- AZA+atRA-induced gene expression program is distinct from spontaneous dormancy
- AZA+atRA reprogramming induces dormancy by enhancing TGF-β-SMAD4 transcriptional program
- SMAD4 depletion renders DCCs resistant to AZA+atRA reprogramming, firing metastatic growth

Authors

Deepak K. Singh, Saul Carcamo, Eduardo F. Farias, ..., Maria Soledad Sosa, Emily Bernstein, Julio A. Aguirre-Ghiso

Correspondence

julio.aguirre-ghiso@einsteinmed.edu

In brief

Awakening of dormant DCCs leads to lethal metastatic disease. Singh et al. show that neo-adjuvant + adjuvant treatment of cancer cells with 5-azacytidine followed by retinoic acid suppresses metastasis by reprogramming malignant cells into dormancy via enhanced TGF-β-SMAD4 signaling.

Singh et al., 2023, Cell Reports *42*, 112560
June 27, 2023
https://doi.org/10.1016/j.celrep.2023.112560

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

5-Azacytidine- and retinoic-acid-induced reprogramming of DCCs into dormancy suppresses metastasis via restored TGF-β-SMAD4 signaling

Deepak K. Singh,[1,2,5,6] Saul Carcamo,[2,3,4] Eduardo F. Farias,[1] Dan Hasson,[2,3,4] Wei Zheng,[5,6] Dan Sun,[1,2,5,6] Xin Huang,[5,6] Julie Cheung,[1,2] Ana Rita Nobre,[1,2] Nupura Kale,[7] Maria Soledad Sosa,[7,12] Emily Bernstein,[2,3] and Julio A. Aguirre-Ghiso[1,2,5,6,8,9,10,11,13,*]

[1]Division of Hematology and Oncology, Department of Medicine and Department of Otolaryngology, Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[2]Department of Oncological Sciences, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[3]Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[4]Bioinformatics for Next Generation Sequencing Facility, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[5]Department of Cell Biology, Albert Einstein College of Medicine, Bronx, NY, USA
[6]Cancer Dormancy and Tumor Microenvironment Institute, Albert Einstein College of Medicine, Bronx, NY, USA
[7]Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[8]Montefiore Einstein Cancer Center, Albert Einstein College of Medicine, Bronx, NY, USA
[9]Gruss-Lipper Biophotonics Center, Albert Einstein College of Medicine, Bronx, NY, USA
[10]Ruth L. and David S. Gottesman Institute for Stem Cell Research and Regenerative Medicine, Albert Einstein College of Medicine, Bronx, NY, USA
[11]Institute for Aging Research, Albert Einstein College of Medicine, Bronx, NY, USA
[12]Present address: Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY, USA
[13]Lead contact
*Correspondence: julio.aguirre-ghiso@einsteinmed.edu
https://doi.org/10.1016/j.celrep.2023.112560

SUMMARY

Disseminated cancer cells (DCCs) in secondary organs can remain dormant for years to decades before re-activating into overt metastasis. Microenvironmental signals leading to cancer cell chromatin remodeling and transcriptional reprogramming appear to control onset and escape from dormancy. Here, we reveal that the therapeutic combination of the DNA methylation inhibitor 5-azacytidine (AZA) and the retinoic acid receptor ligands all-*trans* retinoic acid (atRA) or AM80, an RARα-specific agonist, promotes stable dormancy in cancer cells. Treatment of head and neck squamous cell carcinoma (HNSCC) or breast cancer cells with AZA+atRA induces a SMAD2/3/4-dependent transcriptional program that restores transforming growth factor β (TGF-β)-signaling and anti-proliferative function. Significantly, either combination, AZA+atRA or AZA+AM80, strongly suppresses HNSCC lung metastasis formation by inducing and maintaining solitary DCCs in a $SMAD4^{+}/NR2F1^{+}$ non-proliferative state. Notably, SMAD4 knockdown is sufficient to drive resistance to AZA+atRA-induced dormancy. We conclude that therapeutic doses of AZA and RAR agonists may induce and/or maintain dormancy and significantly limit metastasis development.

INTRODUCTION

Metastasis is a multi-step process and is the primary cause of cancer-related deaths.[1,2] The progression of metastasis is highly complex and depends on the cancer cells ("seeds") and the microenvironment ("soil") of secondary organs.[3,4] Until recently, it was believed that cancer cells disseminated from advanced and clinically aggressive tumor lesions. However, several studies have shown that disseminated cancer cells (DCCs) can be present in secondary organs even before the primary tumor can be detected clinically.[5–8] Importantly, genetically heterogeneous DCC populations appear to remain dormant for long periods (sometimes decades) and have the potential to reactivate as a consequence of additional genetic, chromatin remodeling, and microenvironmental changes.[9–11] These findings suggest that strategies that manage post-extravasation events, and thus DCC biology, may be an effective path to changing patient outcomes by preventing or significantly delaying metastatic overgrowth.

DCCs appear to spend long periods in a quiescent state during clinical dormancy, which is evident by the upregulation of cyclin-dependent kinase inhibitors, like p21 and p27.[9–11] However, many studies have revealed a hierarchy of signaling mechanisms that control dormancy where specific cues such as retinoic acid, BMP4/7, LIF, GAS6, and transforming growth factor β2 (TGF-β2) can activate specific transcription factors (TFs) (e.g., NR2F1,

RARβ, DEC2, NDRG1) that commit cancer cells to a long-lived cell-cycle arrest accompanied by cell plasticity or "stem-like" programs (e.g., SOX9, SOX2, ZFP281).[9–12] The signaling and transcriptional programs mentioned above involve genes controlling proliferation and quiescence, stress tolerance, survival, and pluripotency/self-renewal.[13–15] In addition, dormant cells display a repressive chromatin state that seems long lived, dependent on microenvironmental cues such as retinoic acid, but is also reversible.[15] In fact, we recently demonstrated a role for the repressive macroH2A2 histone variant in suppressing metastasis by tapping into a specific set of genes found in senescent and dormant cells.[16] Taken together, dormant DCCs appear to undergo transitional chromatin and transcriptional reprogramming driven by niche-derived cues.

Previously, we demonstrated the reprogrammable nature of dormancy and showed that the retinoic-acid-regulated nuclear receptor NR2F1 was responsible for inducing and maintaining DCC dormancy. We further showed that 5-azacytidine (AZA) in combination with all-*trans* retinoic acid (atRA) could reprogram tumor cells and led to the induction of quiescence and suppressed head and neck squamous cell carcinoma (HNSCC) primary tumor PDX (patient-derived xenograft) growth *in vivo* in part via NR2F1 function.[15] atRA is the biologically active form of vitamin A, and it mediates its action through RAR and RXR heterodimers.[17] Quiescence and differentiation during development and tissue homeostasis are also regulated by atRA. For example, hematopoietic stem cells respond to atRA by entering a cell-cycle arrest and maintaining a dormant state.[18] Our previous mechanistic studies showed that the TF NR2F1 and a handful of dormancy genes were required for AZA+atRA reprogramming. However, we had not uncovered the extent and functional characteristics of the programs activated by this reprogramming protocol that are NR2F1 independent and whether this strategy could indeed suppress metastatic progression *in vivo*. Mechanistic understanding of AZA+atRA reprogramming is vital because this strategy[15] repurposes two FDA-approved drugs (e.g., AZA and atRA) to treat patients with prostate cancer at risk of developing metastasis (ClinicalTrials.gov: NCT03572387).

Here, we reveal that the AZA+atRA reprogramming protocol only taps into a subprogram of genes found in spontaneously dormant cells (NR2F1 dependent) and instead is dominated by TGF-β-SMAD4-dependent mechanisms (NR2F1 independent or NR2F1 complementary). Importantly, we show that, indeed, AZA+atRA or AZA+AM80 (a clinically approved RARα agonist used in hematological malignancies treatment)[19] used in a neo-adjuvant + adjuvant setting can significantly suppress aggressive HNSCC PDX metastasis to mouse lungs. This phenotype was associated with the induction of NR2F1 and the NR2F1-independent upregulation of TGF-β-SMAD4 expression in DCCs. We further showed that AZA+atRA can also maintain pre-existing dormant DCCs in a further sustained growth-arrested state. Our findings provide insight into the mechanisms behind the reprogramming of cancer cells using AZA+atRA, and identify additional markers (e.g., SMAD4) to pinpoint spontaneously occurring dormant cancer cells or DCCs in response to epigenetic therapies. Our data also support that even cancer cells with highly aberrant genomes can be reprogrammed into a dormant phenotype that curtails metastatic progression.

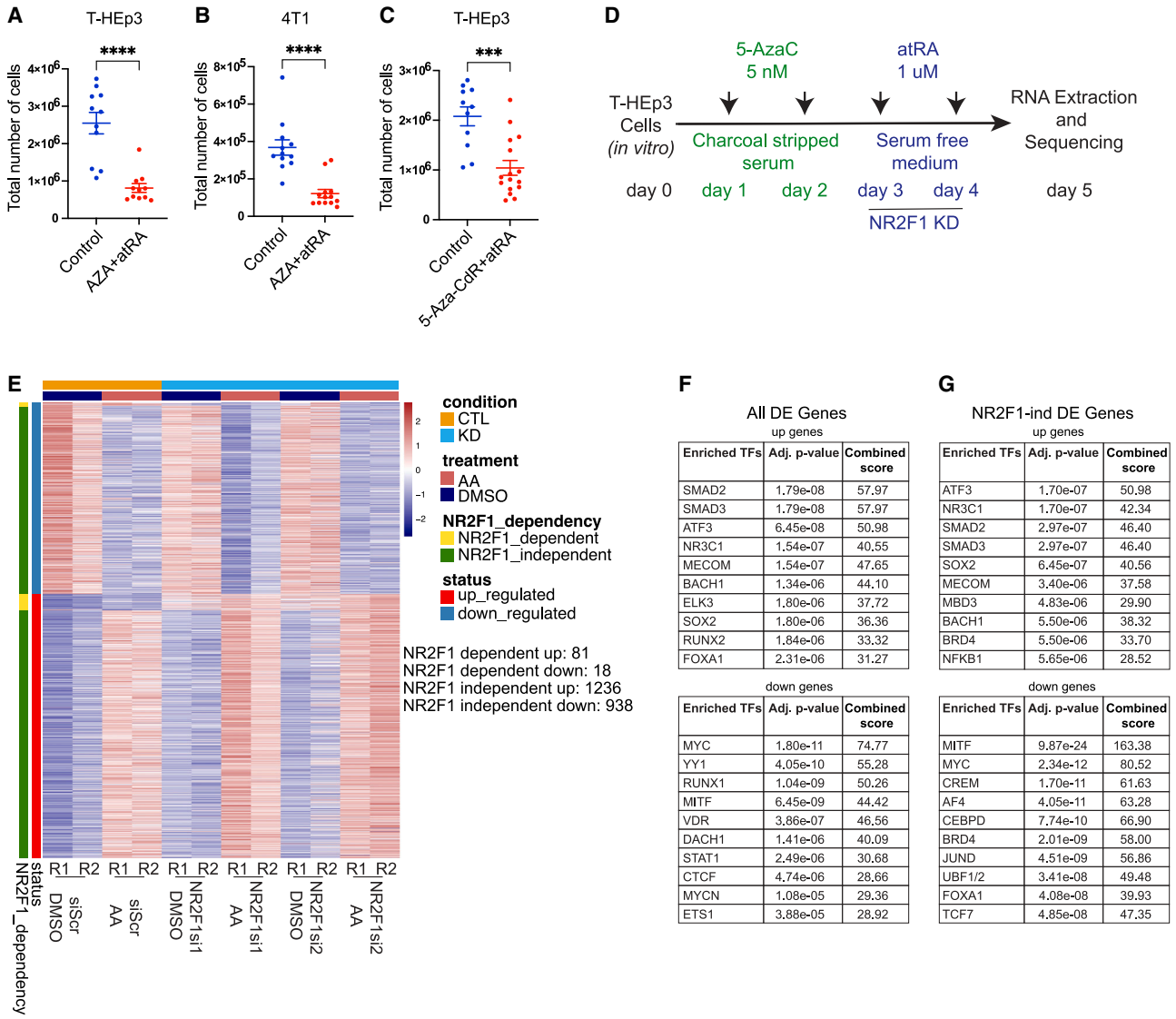
RESULTS

Transcriptional programs controlled by AZA+atRA reprogramming of cancer cells

In this study, we used a combination of HNSCC and breast cancer models from human and mouse origin. To study spontaneously solitary DCCs and conversions to metastasis that occur at high frequency, we employed an *in-vivo*-maintained HNSCC PDX line that is tumorigenic and metastatic to lymph node and lung (T-HEp3) with high efficiency.[20] T-HEp3 cells were originally obtained from a lymph node metastasis with primary carcinoma in the buccal mucosa and were continuously maintained *in vivo*. Transplantation of T-HEp3 cells leads to primary tumor development with 100% efficiency at 7–10 days.[20] After surgery of primary tumors, mice followed for 4–8 weeks will develop metastasis with >80% efficiency.[20] Importantly, even when tumors are small (100–200 mm^3), solitary dormant DCCs can be detected in lymph nodes, lungs, and bone marrow; in the latter organ, they do not form overt metastasis.[21] Thus, unlike in *in-vitro*-maintained cell lines, cancer cells from this *in-vivo*-maintained line allow for efficient study of spontaneous metastasis development. It was also reported that when T-HEp3 cells are maintained over 40 generations *in vitro*, they reprogram in a non-clonal fashion, and when implanted *in vivo*, they enter a dormant state where 100% of implanted cells enter G0/G1 arrest within 48 h; this *in-vitro*-maintained line is termed dormant or D-HEp3.[22,23] This dormant state can last for >16 weeks, and during this time, D-HEp3 cells can spontaneously revert into growth, providing a model to study the reversibility of malignancy. This model has been used over many studies to unravel mechanisms of dormancy and metastasis.[15,16,21,24–27] We also used the 4T1 mouse triple-negative breast cancer model that maintains tumorigenic and metastatic potential.[28]

With these models in hand, we sought to explore the global gene expression programs and pathways regulated by AZA+atRA-mediated reprogramming and to improve our understanding on how to use AZA+atRA or other epigenetic therapies for metastasis suppression. To this end, we first confirmed the growth-suppressive effect of AZA+atRA in malignant T-HEp3 (human HNSCC) and 4T1 (mouse triple-negative breast cancer) cells (Figures 1A and 1B). AZA+atRA reprogramming of T-HEp3 and 4T1 cells significantly inhibited their growth *in vivo* in chorioallantoic membrane (CAM) assays, indicating that at least these two epithelial cancers respond to AZA+atRA treatment with growth suppression. Since AZA also incorporates into RNA molecules and could have other functions, we also tested decitabine or 5-AZA-CdR (5-aza-2′-deoxycytidine), which mainly binds to DNA.[29] A similar growth-suppressive effect was observed with 5-AZA-CdR+atRA (Figure 1C). We also tested whether a histone deacetylase inhibitor (MS275) followed by atRA could generate similar growth suppression, but we did not observe any growth-inhibitory effect (Figure S1A). As atRA alone can have a reversible growth-suppressive effect,[30] it is possible that MS275 may even block atRA signaling.

Next, we explored how AZA+atRA reprogramming impacts the transcriptional landscape of T-HEp3 HNSCC malignant cells compared with the programs activated in spontaneously dormant D-HEp3 cells.[22] Further, we also sought to identify the



F

All DE Genes

up genes

Enriched TFs	Adj. p-value	Combined score
SMAD2	1.79e-08	57.97
SMAD3	1.79e-08	57.97
ATF3	6.45e-08	50.98
NR3C1	1.54e-07	40.55
MECOM	1.54e-07	47.65
BACH1	1.34e-06	44.10
ELK3	1.80e-06	37.72
SOX2	1.80e-06	36.36
RUNX2	1.84e-06	33.32
FOXA1	2.31e-06	31.27

down genes

Enriched TFs	Adj. p-value	Combined score
MYC	1.80e-11	74.77
YY1	4.05e-10	55.28
RUNX1	1.04e-09	50.26
MITF	6.45e-09	44.42
VDR	3.86e-07	46.56
DACH1	1.41e-06	40.09
STAT1	2.49e-06	30.68
CTCF	4.74e-06	28.66
MYCN	1.08e-05	29.36
ETS1	3.88e-05	28.92

G

NR2F1-ind DE Genes

up genes

Enriched TFs	Adj. p-value	Combined score
ATF3	1.70e-07	50.98
NR3C1	1.70e-07	42.34
SMAD2	2.97e-07	46.40
SMAD3	2.97e-07	46.40
SOX2	6.45e-07	40.56
MECOM	3.40e-06	37.58
MBD3	4.83e-06	29.90
BACH1	5.50e-06	38.32
BRD4	5.50e-06	33.70
NFKB1	5.65e-06	28.52

down genes

Enriched TFs	Adj. p-value	Combined score
MITF	9.87e-24	163.38
MYC	2.34e-12	80.52
CREM	1.70e-11	61.63
AF4	4.05e-11	63.28
CEBPD	7.74e-10	66.90
BRD4	2.01e-09	58.00
JUND	4.51e-09	56.86
UBF1/2	3.41e-08	49.48
FOXA1	4.08e-08	39.93
TCF7	4.85e-08	47.35

Figure 1. Transcriptional changes induced by AZA+atRA reprogramming

(A) T-HEp3 cells reprogrammed with AZA+atRA followed by CAM assay.

(B) 4T1, a breast cancer cell line reprogrammed with AZA+atRA followed by CAM assay.

(C) T-HEp3 cells reprogrammed with AZA-CdR+atRA followed by CAM assay. All CAM assays were repeated at least three times.

(D) AZA+atRA reprogramming protocol: T-HEp3 cells seeded at low density were treated with AZA (5 nM) in DMEM containing charcoal-stripped FBS for 2 days at an interval of 24 h. After 48 h, AZA-containing culture medium was replaced with serum-free DMEM, and cells were treated with atRA (1 μM) and NR2F1 knockdown was carried out simultaneously for 48 h. Cells were collected and processed for RNA extraction, followed by RNA sequencing.

(E) A heatmap shows DEGs upon AZA+atRA reprogramming and genes dependent or independent of NR2F1.

(F) ChIP enrichment analysis of DEGs in (E) using Enrichr to compute overrepresentation of transcription factor (TF) targets. Only a few of the top significantly enriched and study-relevant TFs are shown.

(G) ChEA of DEGs in AZA+atRA reprogramming independent of NR2F1 using Enrichr to compute overrepresentation of TF targets.

See also Figures S1 and S2.

NR2F1-dependent and NR2F1-independent/-complementary gene expression programs controlled by AZA+atRA. We treated T-HEp3 primary culture cells from growing PDX tumors, depleted of NR2F1 by specific siRNAs, sequentially with AZA (5 nM) and atRA (1 μM) and performed RNA sequencing (RNA-seq) (Figures 1D and S1B). Principal-component analysis (PCA) of the RNA-seq data across samples showed that replicate samples for each condition are highly similar (Figure S1C). Following this, we performed differential gene expression analysis between multiple conditions, i.e., control vs. reprogrammed and scrambled siRNA (siScr) vs. NR2F1 (si1 and si2 [log2 fold change (FC) < or > 0, adjusted p value [p-adj] < 0.05) (Figure 1E). There were 1,317 genes upregulated and 956 genes downregulated by AZA+atRA reprogramming compared with DMSO control in

the siScr group (Table S1). Several well-known targets of atRA signaling (STRA6, RARβ, HOXA5, and TIMP1) were upregulated[31–33] and were validated by qPCR (Figure S1D), confirming that AZA+atRA-mediated reprogramming in T-HEp3 cells was effective for those genes. Pathway enrichment analysis by Enrichr showed enrichment of epithelial-to-mesenchymal transition (EMT), hypoxia, unfolded protein response, and TGF-β signaling pathways in genes upregulated by AZA+atRA reprogramming (Figure S1E). Accordingly, studies by us and others showed that the TGF-β/SMAD pathway plays an important role in normal stem cell quiescence and cancer cell dormancy, where a higher ratio of TGF-β2/TGF-β1 promotes dormancy via the TGF-βRIII receptor.[21,34–36] We next performed chromatin immunoprecipitation (ChIP) enrichment analysis (ChEA)[37] using DEGs upon AZA+atRA reprogramming and found a significant enrichment of SMAD2 and SMAD3 (adj p = 1.79e−08) target genes in the upregulated DEGs and enrichment of MYC target genes in the downregulated genes (Figure 1F; Table S1). This suggests that AZA+atRA induces a SMAD-dependent program of growth arrest that may be associated with restored or enhanced TGF-β2 induction[15] and suppression of MYC programs.

AZA+atRA reprogramming drives an NR2F1-independent or -complementary transcriptional program

Since NR2F1 is functionally linked to dormancy *in vivo*, and NR2F1 knockdown in D-HEp3 and subsequent *in vivo* CAM assay led to a rapid switch from dormancy to the proliferative stage[15] (Figure S2A), we investigated the NR2F1-dependent gene expression programs activated by AZA+atRA reprogramming. We analyzed the RNA-seq data from T-HEp3 cells depleted of NR2F1 by small interfering RNAs (siRNAs) and reprogrammed them with AZA+atRA (Figures 1D and 1E). To our surprise, differential gene expression analysis showed a small set of genes, 81 genes upregulated and 18 genes downregulated, that were dependent on NR2F1 in AZA+atRA reprogramming (Figure 1E; Table S1). The small number of NR2F1-dependent (NR2F1-dep) DEGs have not provided any robust enriched pathways by AZA+atRA reprogramming in culture. Nevertheless, their contribution must be significant given that NR2F1 loss of function abrogates AZA+atRA reprogramming.[15]

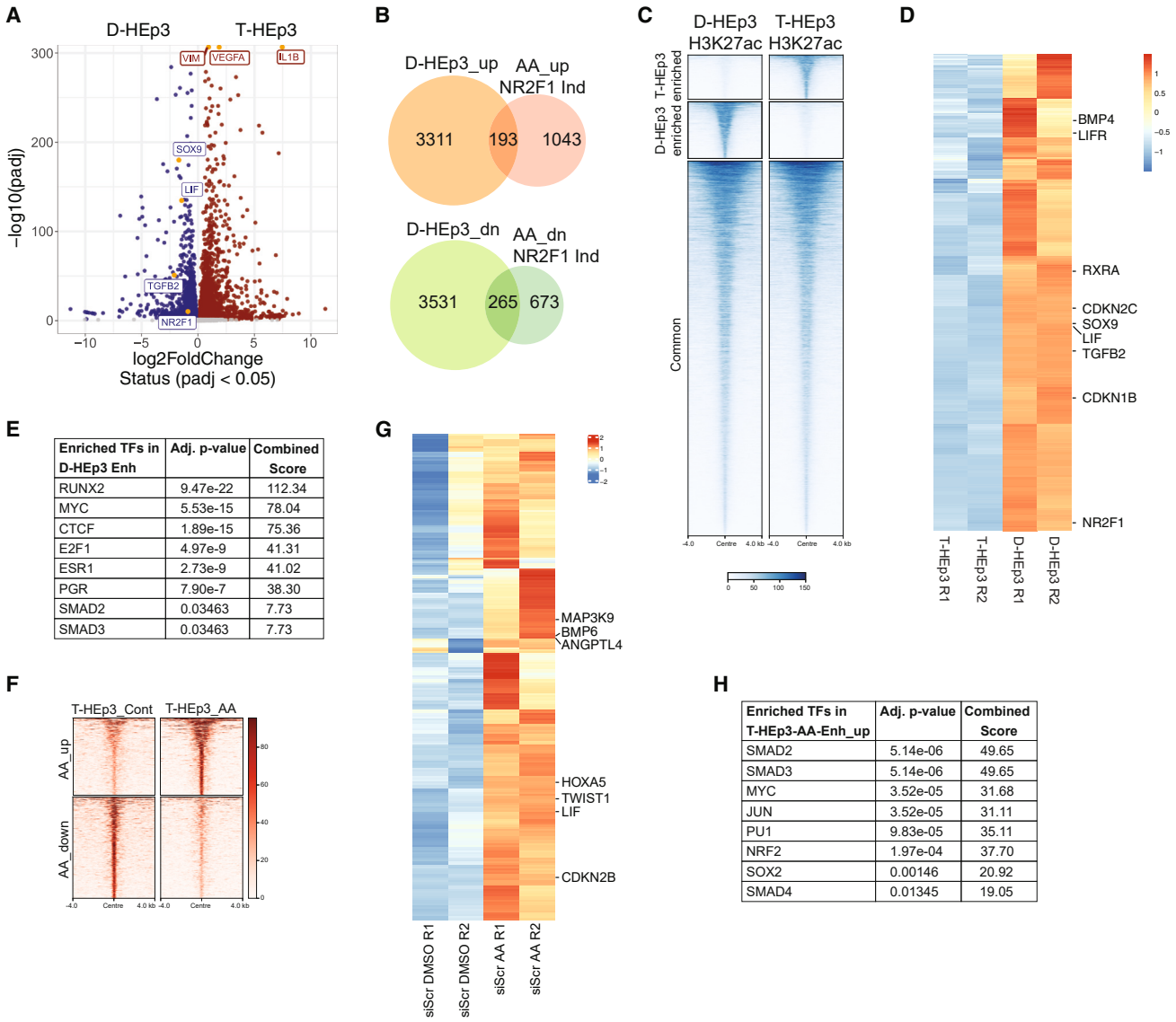
Next, we examined the NR2F1-independent genes responsive to AZA+atRA treatment. Here, we found 938 downregulated and 1,236 upregulated DEGs in AZA+atRA reprogramming, which were independent of NR2F1 expression (Figure 1E; Table S1). This implies that while NR2F1 is important for the reprogramming and spontaneous dormancy,[15] a larger program(s) is activated independently of NR2F1 by this therapeutic approach, but it could complement the NR2F1 function parallelly. Pathway analysis by Enrichr showed that the significantly enriched hallmark gene sets upregulated by AZA+atRA reprogramming in an NR2F1-independent (NR2F1-ind) manner were EMT, hypoxia, TGF-β signaling, and KRAS signaling, while downregulated gene sets were enriched for MYC target and G2M checkpoints (Figure S2B; Table S1). ChEA of the DEGs in AZA+atRA reprogramming that are NR2F1-ind again revealed enrichment of SMAD2, SMAD3, and SOX2 TFs (Figure 1G; Table S1). These data support that, as hinted at by the induction of TGF-β2 by AZA+atRA,[15] activation of TGF-β-SMAD signaling may be important for NR2F1-ind or -complementary AZA+atRA-induced dormancy.

AZA+atRA induces a dormancy-like gene expression program that is distinct from spontaneous dormancy

To compare the similarity in gene expression signatures between the dormancy induced by AZA+atRA in T-HEp3 and spontaneous dormancy programs in D-HEp3 cells,[22] we performed RNA-seq comparing T-HEp3 and D-HEp3 cells. *In vivo* proliferative T-HEp3 primary culture cells at passage 0 (P0) and *in vivo* dormancy D-HEp3 at passages >130[22] were used. Analysis of the RNA-seq data revealed that 3,801 and 3,511 genes were differentially regulated in T-HEp3 and D-HEp3, respectively (Figure 2A). Confirming the phenotypes, some of the critical genes previously found upregulated in D-HEp3 cells and described by microarray data[15] were also found in the RNA-seq, including TGF-β2, NR2F1, SOX9, LIF, and KLF4, while in T-HEp3 cells, MMPs (1, 2, 3, 8, 9, 13, 17, and 19), VIM, VEGFA, interleukin-1B (IL-1B), IL-24, and JUN were expressed (Figure 2A; Table S2). Pathway analysis of DEGs showed that genes upregulated in D-HEp3 were enriched for hallmark pathway E2F targets, unfolded protein response (UPR), and hypoxia, and upregulated genes in T-HEp3 cells were enriched for hallmark pathways such as tumor necrosis factor α (TNF-α) signaling and EMT (Figure S3A; Table S2). Since AZA+atRA-reprogrammed T-HEp3 cells were growth suppressed (Figure 1A), we compared the DEGs of AZA+atRA reprogramming with the DEGs of the D-HEp3 cell line. We observed that 209 (15.86%) upregulated and 271 (28.34%) downregulated genes (of all DEGs) of AZA+atRA-reprogrammed T-HEp3 cells overlapped with upregulated and downregulated genes, respectively, in spontaneously dormant D-HEp3 cells (Figure S3B; Table S3). Next, we checked the NR2F1-ind DEGs of AZA+atRA reprogramming and those differentially expressed in D-HEp3. A total of 458 (21.06%) DEGs of AZA+atRA reprogramming overlap with DEGs of D-HEp3, which accounts for 193 (15.61%) upregulated and 265 (28.25%) downregulated genes (Figure 2B; Table S3). Some of the genes that showed a positive correlation between dormant D-HEp3 cells and AZA+atRA-reprogrammed T-HEp3 cells (NR2F1-ind) were LIF, MAP3K9, CDH4, FOXQ1, MYC, and IL-24 (Table S3). However, the data revealed that ∼80% of the genes induced and repressed by AZA+atRA reprogramming of T-HEp3 cells did not overlap with those regulated in D-HEp3 cells, arguing for a specific program activated by this drug and morphogen combination. Thus, the AZA+atRA reprogramming strategy (both NR2F1-dep and NR2F1-ind), which is effective in generating a long-lived growth suppression, only taps into a subprogram found in spontaneously dormant D-HEp3 cells. The data also revealed that other AZA+atRA specific gene programs may also be able to cooperate with the changes observed in 20% of genes in D-HEp3 cells to induce a strong and long-lived growth suppression.

Epigenetic reprogramming by AZA+atRA involves the SMAD pathway

To further explore whether TGF-β-SMAD signaling could be a pathway activated epigenetically, we focused on enhancer



E

Enriched TFs in D-HEp3 Enh	Adj. p-value	Combined Score
RUNX2	9.47e-22	112.34
MYC	5.53e-15	78.04
CTCF	1.89e-15	75.36
E2F1	4.97e-9	41.31
ESR1	2.73e-9	41.02
PGR	7.90e-7	38.30
SMAD2	0.03463	7.73
SMAD3	0.03463	7.73

H

Enriched TFs in T-HEp3-AA-Enh_up	Adj. p-value	Combined Score
SMAD2	5.14e-06	49.65
SMAD3	5.14e-06	49.65
MYC	3.52e-05	31.68
JUN	3.52e-05	31.11
PU1	9.83e-05	35.11
NRF2	1.97e-04	37.70
SOX2	0.00146	20.92
SMAD4	0.01345	19.05

Figure 2. RNA-seq analysis of D-HEp3 and T-HEp3 transcriptome

(A) Volcano plot showing the DEGs (upregulated) in D-HEp3 (blue) and T-HEp3 (red). Adjusted p value ($\leq$0.05) was used as a cutoff to identify DEGs.

(B) Venn diagram shows the comparison of upregulated and downregulated genes in D-HEp3 with up- and downregulated NR2F1-independent genes upon AZA+atRA reprogramming (p < 3.542e−05, top, and p < 3.975e−07, bottom).

(C) ChIP-seq analysis of H3K27ac marks was performed on D-HEp3 and T-HEp3 cells. Enriched peaks (specific to either D-HEp3 or T-HEp3 and common to both D-HEp3 and T-HEp3) are shown as density plot.

(D) DEGs in D-HEp3 were overlaid with H3K27ac enriched marks in D-HEp3, and the heatmap shows upregulated genes in D-HEp3 regulated by enhancers.

(E) ChEA of enhancer regulated genes in D-HEp3 using Enrichr to compute overrepresentation of TF targets.

(F) ChIP-seq analysis of H3K27ac marks in T-HEp3 control and AZA+atRA reprogrammed.

(G) Upregulated genes in T-HEp3 reprogrammed with AZA+atRA were overlaid with H3K27ac enriched marks, and the heatmap shows the genes regulated by enhancer upon AZA+atRA reprogramming.

(H) ChEA of enhancer regulated genes in T-HEp3 cells reprogrammed with AZA+atRA using Enrichr to compute overrepresentation of TF targets.

See also Figures S3 and S4.

activity. Changes in enhancer activity play an important role in metastatic progression.[38,39] The activity of enhancers is defined by specific chromatin signatures like histone modifications (H3K4me1 and H3K27ac) and chromatin accessibility.[40] To this end, we conducted ChIP coupled to high-throughput sequencing (ChIP-seq) for H3K27ac as a marker of active enhancers in T- and D-HEp3 cells.[41] We identified 2,296 and 2,950 differential enhancers in T- and D-HEp3 cells, respectively (Figure 2C; Table S4). Next, differentially enriched distal/intragenic H3K27ac regions were associated with the positively correlated promoters of DEGs in T-HEp3 vs. D-HEp3 cells within a genomic region of ±500 kb, and we identified 1,220 and 1,035

genes regulated by enhancers, respectively (Figures 2D and S4A; Table S4). This analysis revealed that dormancy-associated genes upregulated in D-HEp3, such as TGF-β2, NR2F1, SOX9, and LIF, were indeed associated with differential enhancers. ChEA of DEGs in D-HEp3 (Figure S4B; Table S2) as well as genes regulated by enhancers in D-HEp3 showed a significant enrichment of SMAD2 and SMAD3 (Figure 2E; Table S4). Next, we looked for the genes differentially expressed (all and NR2F1-ind) by AZA+atRA and regulated by enhancers and found 59 (in all upregulated) and 56 (in NR2F1-ind upregulated) genes regulated by enhancers (Figures S4C and S4D; Table S5). This suggests that the majority of enhancer-regulated genes in Aza+atRA reprogramming are NR2F1-ind. Considering the fact that only a small fraction of AZA+atRA upregulated genes (5.96%) overlap with DEGs in D-HEp3, the number of genes (4.47% all-up and 4.53% NR2F1-ind-up) regulated by enhancers was appreciable (Figure S4E).

In order to identify the genes regulated by enhancers upon AZA+atRA reprogramming, we performed H3K27ac ChIP-seq analysis in T-HEp3 cells reprogrammed with AZA+atRA. We identified 10,346 peaks enriched in T-HEp3 control and reprogrammed with AZA+atRA. Totals of 437 and 573 enhancers were upregulated and downregulated, respectively, in AZA+atRA reprogrammed T-HEp3 (Figures 2F and S4F; Table S6). Next, differentially enriched distal/intragenic H3K27ac regions were associated with the promoters of DEGs in T-HEp3 siScr AZA+atRA within a genomic region of ±500 kb, and we identified 266 (upregulated) and 287 (downregulated) genes regulated by enhancers (Figures 2G and S4G; Table S6). ChEA of upregulated genes upon AZA+atRA reprogramming and regulated by enhancers also showed enrichment of SMAD2/3 and SMAD4 TFs (Figure 2H; Table S6). Further, we compared differentially enriched enhancers upon AZA+atRA reprogramming with the enhancers enriched in D-HEp3 and T-HEp3. Enhancers upregulated upon reprogramming were enriched in D-HEp3, but downregulated enhancers did not show considerable difference between D-HEp3 and T-HEp3 enhancers (Figure S4H; Table S6). This suggests and further highlights that the differences in transcriptional programs in AZA+atRA-reprogrammed T-HEp3 and D-HEp3 could be due to differentially regulated enhancers. Our analysis showed enrichment of SMAD2 and SMAD3 target genes was commonly present in AZA+atRA reprogramming and was enhancer-regulated, suggesting a prominent role of the SMAD pathway in cancer cell dormancy.

Enhanced TGF-β and SMAD4 signaling is a hallmark of AZA+atRA reprogramming into dormancy

The results revealing a SMAD2/3 signaling signature suggested that the AZA+atRA reprogramming strategy might restore to some extent the canonical growth-suppressive function of the TGF-β pathway active during the early stages of cancer progression, which is subsequently lost.[42] To explore this possibility, we first tested the expression of SMAD (2, 3, and 4) proteins in AZA+atRA-reprogrammed T-HEp3 cells. We observed that all three SMADs are, as expected, low in malignant T-HEp3 cells, and AZA+atRA treatment caused an increase in the total level of SMAD2, SMAD3, and SMAD4 (Figure 3A). We previously showed that TGF-β2-induced DCC dormancy in the bone marrow via an activated P-p38 as well as SMAD2 and SMAD1/5 pathway induced expression of p27 leading to quiescence.[34] Upon AZA+atRA reprogramming, we observed an increase in the mRNA levels of TGF-β2, along with TGF-β1 and TGF-β3 (Figure S5A). However, we did not observe any considerable change in the expression or phosphorylation of the SMAD1/5 pathway as reported for TGF-β2[21] (Figure 3A), arguing that AZA+atRA reprogramming may induce a canonical pathway activation. We also documented an increase in the expression of TGF-βRI and TGF-βRIII (Figure S5B), both previously implicated in DCC dormancy in the bone marrow (BM) and lung.[34] Upon AZA+atRA reprogramming, we also detected a strong increase in phosphorylated (P)-p38 and p27, while P-Erk1/2 or total Erk1/2 remained constant, which is similar to our previous observation of a low ERK/p38 activity ratio[15] (Figure 3A). Phosphorylated SMAD3 did not differ between control and AZA+atRA reprogramming, likely because the reprogramming takes days and the pathways reach a steady state, unlike when cells are treated acutely with TGF-β ligands.[21,34]

SMAD2 and SMAD3 proteins frequently shuttle between the cytoplasm and the nucleus, interacting with SMAD4 through the formation of a heteromeric complex, and control transcription of target genes, but the basal levels of SMAD proteins in the cytoplasm and the nucleus usually remain constant.[43,44] Therefore, since we observed an increased expression of SMAD2/3/4, but not phosphorylation of SMAD3, we anticipated their higher nuclear localization would reflect a stronger pathway activation and transcriptional activation of target genes. Further, since phosphorylation of SMAD proteins occurs at early time points and its level fluctuates over time, nuclear translocation of SMAD proteins was the best measure of transcriptional activity in the context of long-term reprogramming treatment. Considering the essential function of SMAD4 in nuclear transport of SMAD2/3 and its role as a tumor suppressor,[45] we reprogrammed T-HEp3 cells with AZA+atRA and performed SMAD4 and p27 immunofluorescence (IF) and immunoblot assays. We observed an increased nuclear localization for SMAD4 and p27 (Figures 3B, 3C, and S5C). Similarly, reprogramming of T-HEp3 cells with AZA+AM80 (50 nM) also led to an increase in SMAD4 and p27 protein levels as detected by immunoblot (Figure S5C–S3E). These data show that AZA+atRA reprogramming orchestrates enhanced TGF-β/SMAD signaling, notably with upregulation of SMAD4, in malignant HNSCC cells.

To determine whether AZA+atRA was activating a TGF-β pathway as observed in dormant D-HEp3 cells, we examined the expression level of SMAD proteins in T- and D-HEp3 cells. We observed higher expression of SMAD2, SMAD3, and SMAD4 in D-HEp3 cells as well (Figure 3D). In addition to higher expression, there was more nuclear-localized SMAD4 in D-HEp3 cells (Figures 3E and 3F). Importantly, knockdown of SMAD4 was sufficient to reverse the AZA+atRA induction of p27 in T-HEp3 cells and partially reversed AZA+atRA-induced growth suppression of T-HEp3 cells in the CAM *in vivo* assay (Figures 3G, 3H, S5F, and S5G). These data support that AZA+atRA reprogramming induces, independently of NR2F1, a dormancy-like phenotype in proliferative cancer cells, in part by activating a TGF-β-p38-SMAD4 pathway. The data also support that AZA+atRA reprogramming is robust in inducing growth

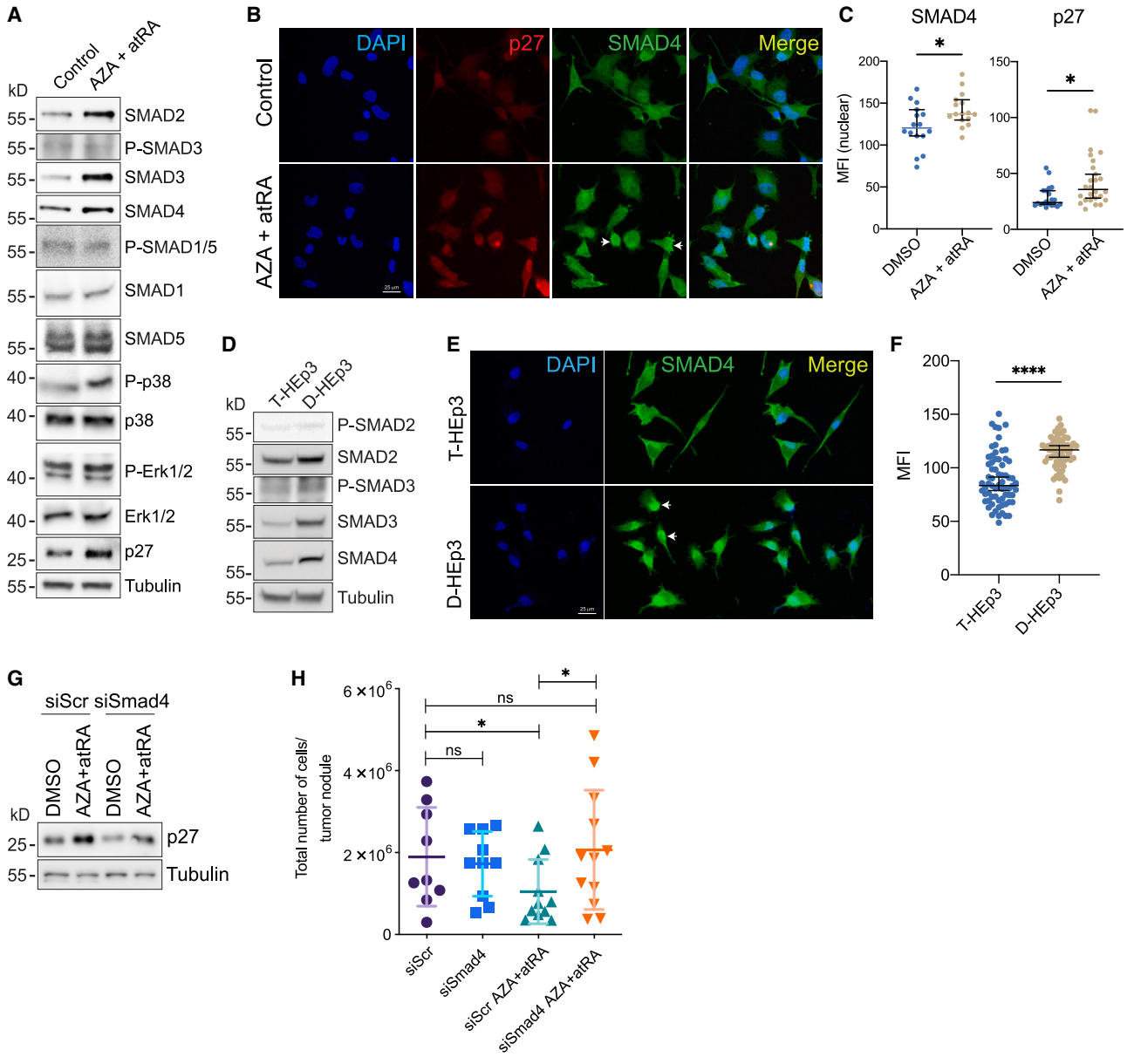


Figure 3. AZA+atRA reprogramming induces SMAD2, SMAD3, and SMAD4 expression

(A) Western blots show upregulation of SMAD2/3/4 and dormancy-associated marker genes like p27 and P-p38 upon AZA+atRA reprogramming in T-HEp3 cells.

(B) Immunofluorescence (IF) staining show the higher nuclear localization of p27 and SMAD4 (marked with solid white arrowheads) in AZA+atRA-reprogrammed T-HEp3 cells.

(C) Quantification of nuclear SMAD4 and p27 in control and AZA+atRA-reprogrammed T-HEp3 cells (unpaired t test, $p \leq 0.05$).

(D) Western blots show higher expression of SMAD2/3/4 in D-HEp3 cells compared with T-HEp3 cells.

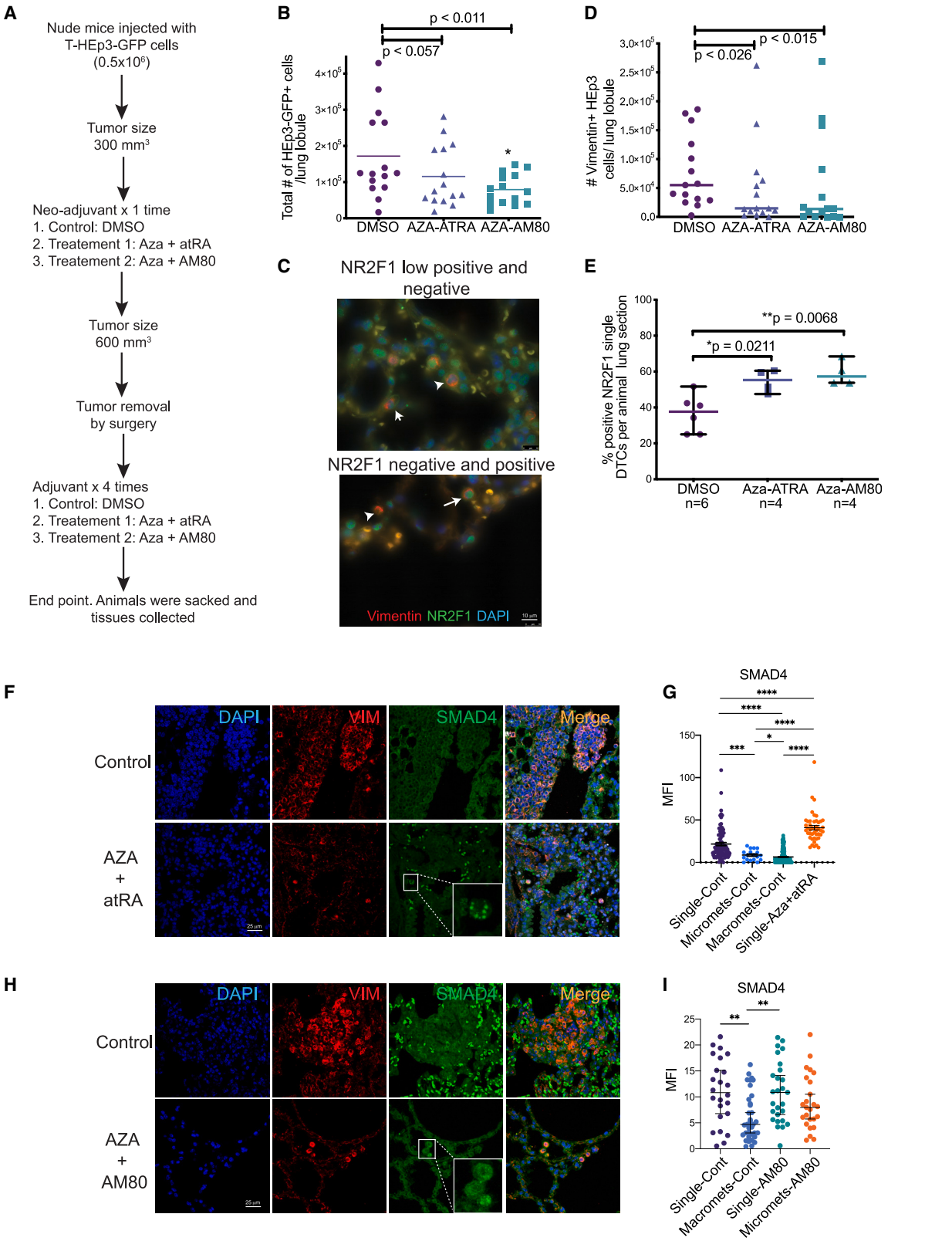
(E) IF staining shows higher SMAD4 nuclear localization (marked with solid white arrowheads) in D-HEp3 cells.

(F) Quantification of nuclear SMAD4 in D-Hep3 and T-HEp3 cells (unpaired t test, $p \leq 0.05$, error bar represents standard error of mean).

(G) SMAD4 knockdown by siRNA and AZA+atRA reprogramming in T-HEp3 cells reveals p27 dependency on SMAD4.

(H) T-HEp3 (1.5×10^{-5}) cells depleted of SMAD4 and AZA+atRA reprogrammed were inoculated in chicken CAM and incubated for 4 days. After incubation, tumors were excised, and the total number of cells was counted (experiment was repeated three times with a minimum of three eggs per experiment per condition, Mann-Whitney test, $p \leq 0.05$, error bar represents standard deviation).

See also Figure S5.



(legend on next page)

suppression because it induces differentiation (RA)- and dormancy-inducing (TGF-β2) pathways.

AZA+atRA or AZA+AM80 induces dormancy of NR2F1+/SMAD4+ DCCs, suppressing metastatic outgrowth in the lungs

Our data support that the epigenetic and transcriptional changes in reprogrammed cancer cells via AZA+atRA treatment, which has only been tested for short periods on tumor growth at the primary site,[15] could induce a long-lived and self-sustained program of canonical TGF-β-SMAD2/3/4 signaling, limiting the metastatic expansion of pre-existing DCCs.

To this end, we tested the effect of AZA treatment followed by atRA or AM80 on metastatic progression. Immunocompromised nude mice (BALB/c nu/nu) were injected subcutaneously (s.c.) with 0.5×10^{-6} T-HEp3-GFP cells obtained directly from the PDX maintained tumors.[46] After primary tumors reached a size of 300 mm^3, mice were treated with vehicle or with one cycle of neoadjuvant (AZA+atRA or AZA+AM80) treatments (Figure 4A). We performed neoadjuvant studies because T-HEp3 cells can disseminate very early after implantation, already seeding lungs when tumors are smaller than 300 mm^3.[21] We used 1 mg/kg/day AZA and atRA and 0.3 mg/kg/day AM80 doses following the previous optimization of the dosing.[47] After the neoadjuvant cycle, tumors (600 mm^3) were removed by surgery, and after 48 h, mice were treated for four adjuvant cycles with the vehicle, AZA+atRA, or AZA+AM80. At the end of the fourth cycle, mice were euthanized, and the number of T-HEp3-GFP$^+$ tumor cells was scored in one lung lobule enzymatically digested with collagenase-I (Figure 4A).

These fresh collagenase-I suspensions were further stained in suspension using anti-vimentin human-specific antibodies as reported previously[15] to provide an independent confirmatory measure. This second quantification was included because we previously reported that dormant DCCs could lower GFP expression, preventing the proper detection of quiescent cells.[21] The other lung lobule was processed for histological and IF analysis for further DCC analysis. Compared with the control vehicle, mice treated with AZA+AM80 showed a significant decrease in the number of T-HEp3-GFP cells in the lungs (Figure 4B). AZA+atRA also showed a decrease in the DCCs but was not as effective as AM80, possibly due to its shorter half-life than AM80.[48] Quantification of the same samples for the number of vimentin$^+$ DCCs confirmed the GFP count differences between groups but revealed that both AZA+atRA- and AZA+AM80-treated mice showed a significant decrease in tumor cell burden in the lungs (Figures 4C and 4D). We observed a lower number of vimentin$^+$ cancer cells than GFP$^+$ cells, potentially due to the loss of cells during washes in the processing for staining. However, the difference between groups remains unchanged. The suppression of metastatic burden correlated with dormancy induction as protein expression of NR2F1 was significantly upregulated in the DCCs and micrometastasis in the lungs of mice treated with AZA+atRA or AZA+AM80 (Figure 4E). We conclude that a 5 week neoadjuvant + adjuvant treatment of mice with already existing metastatic spread can significantly suppress metastatic progression, which correlated with the induction of the dormancy-regulator NR2F1, confirming the activation of the NR2F1-dep program.

Next, we tested whether the SMAD4 program could inform on HNSCC patient outcome. To this end, we tested whether expression of SMAD2/3 and SMAD4 target genes in primary tumors associated with differences in regression-free survival. For the analysis, the SMAD2/3 and SMAD4 target genes were extracted from the DEGs (RNA-seq data of AZA+atRA-reprogrammed T-HEp3) using Enrichr.[49] The regression-free survival analyses were performed with a publicly available RNA-seq data comprised of 124 patients with HNSCC.[50] We found that patients with enrichment for the SMAD2/3 or SMAD4 signature relapsed at a slower rate than patients low for that signature, albeit with limited statistical significance (Figures S6A and S6B). This suggests that DCCs derived from tumors with a high SMAD signature may take longer to activate from dormancy.

Because access to patient's target organs to identify DCCs is limited, we tested this possibility in mice in control or under AZA+atRA treatments. When we tested whether the SMAD4 program could be detected in the mouse lungs treated with vehicle control, AZA+atRA, or AZA+AM80, we found that SMAD4 was upregulated spontaneously in solitary DCCs compared with micro- and macrometastatic lesions that had very low or no expression of SMAD4 (Figures 4F–4I). This argues that spontaneously

Figure 4. AZA+atRA and AZA+AM80 suppress the metastatic load in an experimental mouse model

(A) Schematic of *in vivo* mouse experiment. 0.5×10^{-6} T-HEp3-GFP primary tumor cells (passage 0 from CAM) were injected into nude mice. Mice were treated in a neo-adjuvant setting with AZA+atRA or AZA+AM80 once when the tumor size reached an average size of 300 mm^3. At a 600 mm^3 tumor size, surgery was performed to remove the primary tumors from the animals, and they were treated with four cycles (28 days) of adjuvant therapy (AZA+atRA or AZA+AM80). Animals were euthanized, and organs were processed for analysis.

(B) Mouse lungs digested enzymatically and counted for GFP$^+$ T-HEp3 single cells.

(C) T-HEp3-GFP cells in the digested lung suspensions were stained with vimentin and NR2F1. Representative images show positive (white arrow long tail), low positive (white arrow with short tail), and negative signals (white arrowhead) for NR2F1.

(D) Quantification of T-HEp3-GFP cells in lung sections using vimentin IF staining.

(E) Quantification of NR2F1$^+$ and vimentin$^+$ T-HEp3-GFP cells in lung sections. n = 15 unless otherwise mentioned (Mann-Whitney test, $p \leq 0.05$, error bar represents standard deviation).

(F) IF staining images showing SMAD4 expression in disseminated T-HEp3-GFP cells in control and AZA+atRA-treated mouse lungs.

(G) Quantifications of mean fluorescent intensity (MFI) of SMAD4 signals in T-HEp3-GFP cells in control and AZA+atRA-treated mouse lungs (n from $\geq$5 mice, Mann-Whitney test, $p \leq 0.05$, error bar represents standard deviation).

(H) IF staining images showing SMAD4 expression in disseminated T-HEp3-GFP cells in control and AZA+AM80-treated mouse lungs.

(I) Quantifications of MFI of SMAD4 signals in T-HEp3-GFP cells in control and AZA+AM80-treated mouse lungs (n from $\geq$5 mice, Mann-Whitney test, $p \leq 0.05$, error bar represents standard deviation).

See also Figure S6.

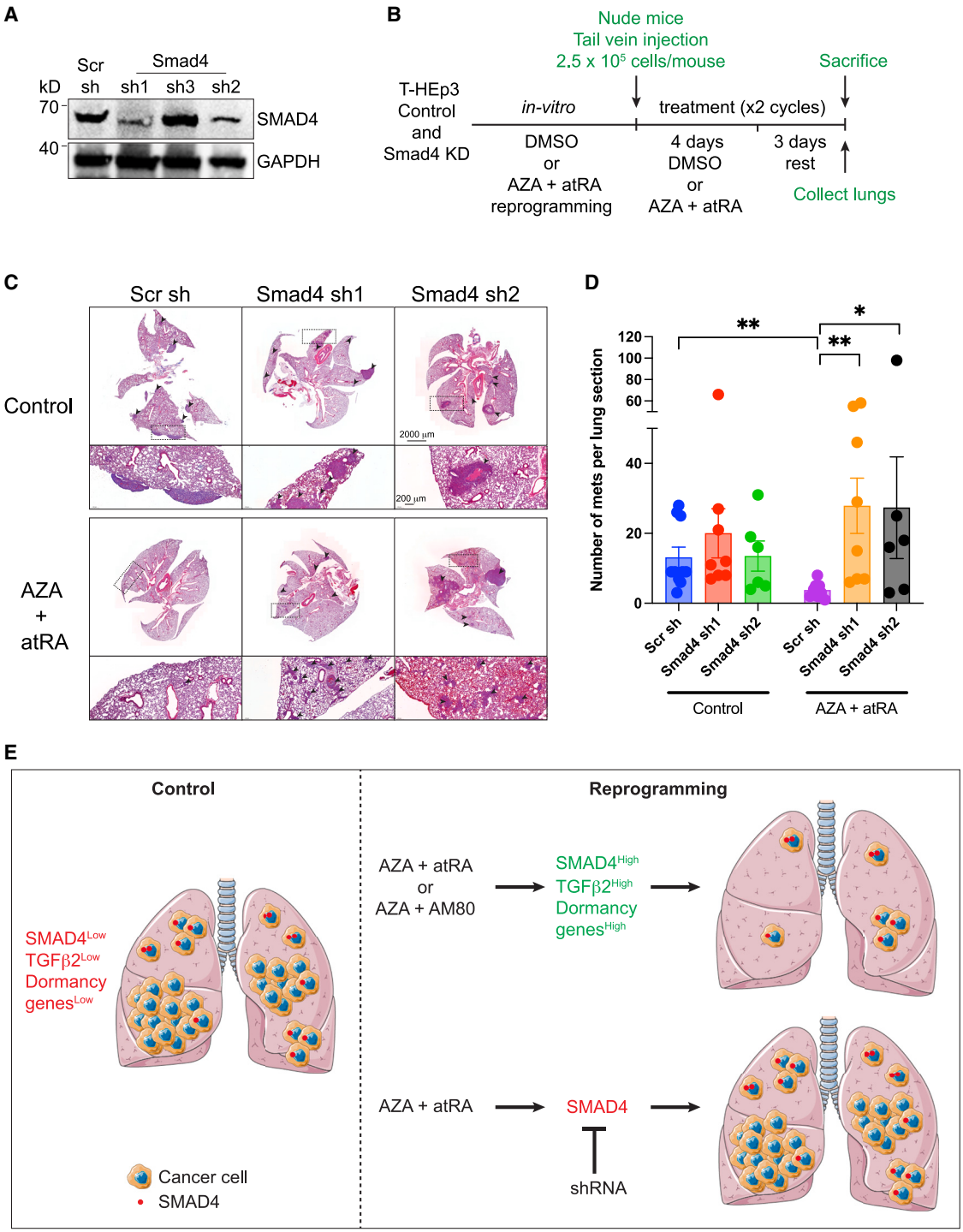


Figure 5. SMAD4 keeps DCCs in a dormant single-cell state

(A) Western blot shows knockdown efficiency of SMAD4 by stably expressed shRNAs in T-HEp3 cell line.

(B) Schematic of *in vivo* mouse experiment. T-HEp3 cells with or without SMAD4 were reprogrammed for a week by AZA+atRA *in vitro* and then injected into nude mice. Mice were treated with AZA (1 mg/kg/day) and atRA (1 mg/kg/day) for 2 weeks. Lungs were collected and processed for experiments.

(C) Representative images of H&E-stained mouse lungs showing metastasis.

(D) Graph shows number of metastatic lesions (mets) per lung section per mouse ($n \geq 6$ animals, Mann-Whitney test, $p \leq 0.05$, error bar represents standard deviation).

(legend continued on next page)

dormant DCCs activate the SMAD4 program. Macrometastases were not detected in the lungs of mice treated with AZA+atRA or AZA+AM80, supporting powerful metastasis-suppressive activity. In the AZA+atRA group, while no micro- or macrometastases were found, even the solitary residual DCCs expressed higher levels of SMAD4. In the AM80 group, solitary DCCs displayed similar levels of SMAD4 to the solitary DCCs in the control group, but the few micrometastases in the AM80 group displayed significant upregulation of SMAD4 compared with metastasis in the control group.

SMAD4 is required for AZA+atRA reprogramming to suppress metastatic outgrowth

Next, to confirm that SMAD4 expression is required to keep DCCs dormant and prevent metastatic proliferation upon AZA+atRA reprogramming, we knocked down SMAD4 and performed reprogramming *in vivo*. To this end, we generated two SMAD4 stable knockdown T-HEp3 cell lines (Figure 5A). Scrambled shRNA and SMAD4 knockdown (sh1 and sh2) cells were reprogrammed with AZA+atRA or DMSO control *in vitro* and then 2.5×10^{-5} cells were injected into the tail vein of nude mice. Mice were treated with AZA (2 days) followed by atRA (2 days) and had 3 days of rest for 2 weeks (Figure 5B). To confirm that the SMAD4 shRNAs were still active (GFP^+), we performed IF to detect SMAD4 in lung sections from mice. We observed a significant loss of SMAD4 in both sh1 and sh2 groups (Figures S7A and S7B). We also observed a significant increase in the expression of SMAD4 in T-HEp3 cells in the lungs of mice treated with AZA+atRA (Figure S7B), confirming our findings in Figure 4. Mice injected with T-HEp3 (scrambled shRNA [Scr-sh]) and treated with DMSO control formed metastatic lesions in the lungs, while those treated with AZA+atRA showed significantly reduced metastatic growth (Figures 5C and 5D). In contrast, upon knockdown of SMAD4 (both sh1 and sh2 groups), AZA+atRA reprogramming was not able to keep cancer cells dormant, and these animals showed strong metastatic proliferation. This demonstrates that AZA+atRA-induced SMAD4 expression is required to keep DCC growth suppressed. Our data support a model where dormant single DCCs express higher levels of SMAD4 (among other pro-dormancy genes) to remain dormant but spontaneous loss of TGF-β-SMAD4 signaling allows for proliferation of DCCs and subsequent metastatic outgrowth. AZA+atRA increases and/or maintains SMAD4 expression across heterogeneous DCCs in different proliferative states by epigenetically reprogramming the cells into dormancy. This reprogramming is associated with enhancer remodeling that keeps a stable dormant single DCC state for long time periods (Figure 5E).

DISCUSSION

This study aimed to better define the gene expression programs and enhancer landscape regulated by the AZA+atRA reprogramming protocol[15] and to test whether such a reprogramming therapy could suppress metastasis via induction of a dormancy-like program. Both goals were achieved, and the data revealed informative differences between therapy-induced and spontaneously activated dormancy pathways.

Analysis of the AZA+atRA-induced transcriptional and enhancer programs showed that reprogramming therapy only regulates a subset and a rather small subprogram found in dormant D-HEp3 cells. However, this program can suppress metastasis development, at least for the time the therapy is administered. Our data also revealed that an even smaller subprogram of gene regulation is dependent on NR2F1 upon AZA+atRA reprogramming of malignant HEp3 cells, which again is important for the growth-suppressive effect of the AZA+atRA treatment. Intriguingly, a TGF-β-SMAD program was evident in the AZA+atRA (AA)-reprogrammed T-HEp3, which was independent of NR2F1 but was activated in parallel to the NR2F1 program. This program was also found in D-HEp3 cells and has been revealed in other models as well.[21] Components of the TGF-β-SMAD program seem to be epigenetically controlled, as our analysis revealed that genes controlled by specific enhancers in D-HEp3 cells also show enrichment for SMAD2/3 regulated genes. This program is associated with TGF-β2 but not TGF-β1 signaling, the latter being crucial for the reactivation of these cancer cells.[21] However, AZA+atRA-induced reprogramming seemed to reverse the pro-growth function of TGF-β1 and restore canonical SMAD4-dependent signaling for this pathway. In fact, TGF-β1 is not a reactivation signal in AZA+atRA-induced dormant T-HEp3 cells. AZA+atRA resulted in the upregulation of all TGF-β ligands and TGF-βRI and TGF-βRIII and strong upregulation of SMAD2, SMAD3, and SMAD4. The latter is commonly silenced or mutated in HNSCC and other cancers,[51,52] and the strong upregulation found in the *in vitro* reprogramming but also in the solitary DCCs in lungs was remarkable. This result and the fact that SMAD4 knockdown (KD) reversed the reprogramming suggested that AZA+atRA may restore, to some extent, the tumor-suppressive function of the TGF-β pathway even in HNSCC PDX tumors with highly aberrant genomes. SMAD4 induction seems to be key for induction of p27 and for growth arrest upon AZA+atRA treatment *in vivo*, as SMAD4 KD induced resistance to AZA+atRA reprogramming. This may also suggest that patients with mutations affecting SMAD4 function may be resistant to AZA+atRA treatment.

The multiplicity of programs, NR2F1-dep and -ind, activated by AZA+atRA treatment may explain the long-lived and durable effect of the reprogramming observed in primary sites,[15] spontaneously induced dormancy of DCCs in the BM of patients with breast cancer,[53] and HNSCC metastasis in this study. Dosing of AZA+atRA using a neo-adjuvant + adjuvant scheme was effective in suppressing metastasis and DCCs upregulated in both NR2F1 and SMAD4, supporting that the programs revealed in the RNA-seq are detected in the growth-arrested DCCs. Notably, the use of these two proteins as biomarkers may help

(E) Working model: lungs of control animals harbor macrometastasis, micrometastasis, and single DCCs, while treatment with AZA+atRA or AZA+AM80 inhibits metastatic growth leading to micrometastasis and single DCCs. Single DCCs have higher SMAD4 expression, which keeps cancer cells in a dormant single-cell state. Knockdown of SMAD4 abolishes the effect of AZA+atRA reprogramming to keep cancer cells dormant.
See also Figure S7.

pinpoint spontaneously dormant DCCs or those induced by epigenetic therapies. Similarly, the use of an NR2F1 and/or SMAD4 signature in primary tumors may identify patients with different risk levels for relapse. Interestingly, the clinically approved agonist for RARα (AM80) was also effective in suppressing metastatic outgrowth after AZA treatment. This may be important for future clinical applications, as the AZA+atRA reprogramming protocol is being used in a prostate cancer clinical trial (ClinicalTrials.gov: NCT03572387), and more potent agonists of RARα may be useful to treat these patients. In addition, we have also found that an NR2F1 agonist can suppress metastasis.[24] Interestingly, a neural crest differentiation program was activated by the NR2F1 agonist,[24] which we did not find with the AZA+atRA protocol. The NR2F1 agonist did not activate the TGF-β pathway,[24] in agreement with our findings in this study that this is an NR2F1-ind or -complimentary mechanism.

Our analysis of enhancers regulated by AZA+atRA revealed an enrichment of SMAD2 and SMAD3 target genes, suggesting a prominent role of the SMAD pathway in the reprogramming-induced dormancy of cancer cells using this epigenetic therapeutic strategy. Further, the presence of active SMAD2/3-linked enhancers in D-HEp3 cells suggests that chromatin remodeling and epigenetic reprogramming may be sufficient to suppress metastatic potential as observed in the transition of T-HEp3 cells to D-HEp3 cells.

Our new data support that therapeutically induced dormant cancer cells may represent an alternative for managing residual latent or progressively growing cancer disease that may not be manageable with standard care approaches. Our studies paired with the clinical trial results may provide important information on how to optimize induced dormant cancer cell therapies in the near future.

Limitations of the study

We cannot rule out the possibility that AZA+atRA-reprogrammed cancer cells depend on other cancer cell-intrinsic mechanisms (i.e., autocrine TGF-β signaling) and/or on an effect on the microenvironment (i.e., effect on endothelial cells, mesenchymal stem cells, myeloid cells, etc.) in addition to the mechanisms proposed here. To explore the effects on the microenvironment, it would be important to use syngeneic systems that allow investigation into how AZA+atRA affects innate and adaptive immune cells as well as non-immune stromal cells that could influence DCC fate. It is also important that the AZA+atRA strategy is tested in additional mouse models and PDX models that faithfully recapitulate spontaneous dissemination. Further studies are needed to determine whether enhancers controlled by AZA+atRA contribute to the stability of the dormancy program and whether the genes they regulate may serve as biomarkers of a durable dormancy phenotype. Additional studies deleting enhancer sequences via CRISPR genome editing are required to determine the functional relevance of such regulatory elements.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112560.

ACKNOWLEDGMENTS

We thank the Aguirre-Ghiso, Bernstein, and Sosa labs for helpful discussions. This work was supported by grants from the National Institutes of Health (NIH)/National Cancer Institute (NCI) (CA109182, CA216248, CA218024, and CA196521 to J.A.A.-G. and RO1CA154683 and CA218024 to E.B.) and The Jimmy V Foundation to J.A.A.-G. J.A.A.-G. is a Samuel Waxman Cancer Research Foundation Investigator. M.S.S. was supported by a Melanoma Research Foundation Career Development Award, NIH/NCI K22CA201054, and a Melanoma Research Alliance team Science Award. RNA-seq and ChIP-seq analyses were supported in part through the Oncological Sciences Sequencing Core supported by the Tisch Cancer Institute of the Icahn School of Medicine at Mount Sinai (ISMMS) Cancer Center Support Grant P30 (CA196521); scientific computing was supported by the Office of Research Infrastructure of the NIH under award number S10OD026880 to ISMMS, the Mount Sinai Genomics Technology Facility, and Deniz Demircioglu from the Bioinformatics for Next Generation Sequencing (BiNGS) shared resource facility within the Tisch Cancer Institute at the Icahn School of Medicine at Mount Sinai. The development of this shared resource is partially supported by the NCI P30 Cancer Center support grant. H and E images were scanned using a 3D Histec Pannoramic 250 Flash II slide scanner (S10OD026852-01A1) at AIF (Cancer Center, P30CA013330), Albert Einstein College of Medicine.

AUTHOR CONTRIBUTIONS

D.K.S. planned and conducted experiments and analyzed data; E.F., D.K.S., J.C., A.R.N., and N.K. performed *in vivo* mouse experiment 1, and D.K.S., W.Z., and X.H. performed *in vivo* mouse experiment 2. D.K.S., S.C., D.H., and D.S. performed RNA-seq and ChIP-seq experiments and analyzed data; and J.A.A.-G., E.B., D.K.S., and M.S.S. analyzed data, provided insight, and wrote the manuscript. J.A.A.-G. conceived the project with M.S.S. and designed experiments along with M.S.S., D.K.S., and E.F. J.A.A.-G., M.S.S., and E.B. secured funding.

DECLARATION OF INTERESTS

J.A.A.-G. is a scientific co-founder of, scientific advisory board member at, and equity owner in HiberCell and receives financial compensation as a consultant for HiberCell, a Mount Sinai spin-off company focused on therapeutics that prevent or delay cancer recurrence.

Received: September 7, 2021
Revised: January 31, 2023
Accepted: May 8, 2023
Published: June 1, 2023

REFERENCES

1. Dillekås, H., Rogers, M.S., and Straume, O. (2019). Are 90% of deaths from cancer caused by metastases? Cancer Med. *8*, 5574–5576. https://doi.org/10.1002/cam4.2474.
2. Lambert, A.W., Pattabiraman, D.R., and Weinberg, R.A. (2017). Emerging biological principles of metastasis. Cell *168*, 670–691. https://doi.org/10.1016/j.cell.2016.11.037.
3. Paget, S. (1889). The distribution of secondary growths in cancer of the breast. Lancet *133*, 571–573. https://doi.org/10.1016/s0140-6736(00)49915-0.
4. Peinado, H., Zhang, H., Matei, I.R., Costa-Silva, B., Hoshino, A., Rodrigues, G., Psaila, B., Kaplan, R.N., Bromberg, J.F., Kang, Y., et al. (2017). Pre-metastatic niches: organ-specific homes for metastases. Nat. Rev. Cancer *17*, 302–317. https://doi.org/10.1038/nrc.2017.6.
5. Sosa, M.S., Bragado, P., and Aguirre-Ghiso, J.A. (2014). Mechanisms of disseminated cancer cell dormancy: an awakening field. Nat. Rev. Cancer *14*, 611–622. https://doi.org/10.1038/nrc3793.
6. Hosseini, H., Obradović, M.M.S., Hoffmann, M., Harper, K.L., Sosa, M.S., Werner-Klein, M., Nanduri, L.K., Werno, C., Ehrl, C., Maneck, M., et al. (2016). Early dissemination seeds metastasis in breast cancer. Nature *540*, 552–558. https://doi.org/10.1038/nature20785.
7. Harper, K.L., Sosa, M.S., Entenberg, D., Hosseini, H., Cheung, J.F., Nobre, R., Avivar-Valderas, A., Nagi, C., Girnius, N., Davis, R.J., et al. (2016). Mechanism of early dissemination and metastasis in Her2(+) mammary cancer. Nature *540*, 588–592. https://doi.org/10.1038/nature20609.
8. Klein, C.A. (2020). Cancer progression and the invisible phase of metastatic colonization. Nat. Rev. Cancer *20*, 681–694. https://doi.org/10.1038/s41568-020-00300-6.
9. Aguirre-Ghiso, J.A., and Sosa, M.S. (2018). Emerging topics on disseminated cancer cell dormancy and the paradigm of metastasis. Annu. Rev. Cell Biol. *2*, 377–393. https://doi.org/10.1146/annurev-cancerbio-030617-050446.
10. Phan, T.G., and Croucher, P.I. (2020). The dormant cancer cell life cycle. Nat. Rev. Cancer *20*, 398–411. https://doi.org/10.1038/s41568-020-0263-0.
11. Singh, D.K., Patel, V.G., Oh, W.K., and Aguirre-Ghiso, J.A. (2021). Prostate cancer dormancy and reactivation in bone marrow. J. Clin. Med. *10*, 2648. https://doi.org/10.3390/jcm10122648.
12. Nobre, A.R., Dalla, E., Yang, J., Huang, X., Wullkopf, L., Risson, E., Razghandi, P., Anton, M.L., Zheng, W., Seoane, J.A., et al. (2022). ZFP281 drives a mesenchymal-like dormancy program in early disseminated breast cancer cells that prevents metastatic outgrowth in the lung. Nat. Can. (Ott.) *3*, 1165–1180. https://doi.org/10.1038/s43018-022-00424-8.
13. Adam, A.P., George, A., Schewe, D., Bragado, P., Iglesias, B.V., Ranganathan, A.C., Kourtidis, A., Conklin, D.S., and Aguirre-Ghiso, J.A. (2009). Computational identification of a p38SAPK-regulated transcription factor network required for tumor cell quiescence. Cancer Res. *69*, 5664–5672. https://doi.org/10.1158/0008-5472.CAN-08-3820.
14. Kim, R.S., Avivar-Valderas, A., Estrada, Y., Bragado, P., Sosa, M.S., Aguirre-Ghiso, J.A., and Segall, J.E. (2012). Dormancy signatures and metastasis in estrogen receptor positive and negative breast cancer. PLoS One *7*, e35569. https://doi.org/10.1371/journal.pone.0035569.
15. Sosa, M.S., Parikh, F., Maia, A.G., Estrada, Y., Bosch, A., Bragado, P., Ekpin, E., George, A., Zheng, Y., Lam, H.M., et al. (2015). NR2F1 controls tumour cell dormancy via SOX9- and RARbeta-driven quiescence programmes. Nat. Commun. *6*, 6170. https://doi.org/10.1038/ncomms7170.
16. Sun, D., Singh, D.K., Carcamo, S., Filipescu, D., Khalil, B., Huang, X., Miles, B.A., Westra, W., Sproll, K.C., Hasson, D., et al. (2022). MacroH2A impedes metastatic growth by enforcing a discrete dormancy program in disseminated cancer cells. Sci. Adv. *8*, eabo0876. https://doi.org/10.1126/sciadv.abo0876.
17. Lefebvre, P., Martin, P.J., Flajollet, S., Dedieu, S., Billaut, X., and Lefebvre, B. (2005). Transcriptional activities of retinoic acid receptors. Vitam. Horm. *70*, 199–264. https://doi.org/10.1016/S0083-6729(05)70007-8.
18. Cabezas-Wallscheid, N., Buettner, F., Sommerkamp, P., Klimmeck, D., Ladel, L., Thalheimer, F.B., Pastor-Flores, D., Roma, L.P., Renders, S., Zeisberger, P., et al. (2017). Vitamin A-retinoic acid signaling regulates hematopoietic stem cell dormancy. Cell *169*, 807–823.e19. https://doi.org/10.1016/j.cell.2017.04.018.
19. Sanda, T., Kuwano, T., Nakao, S., Iida, S., Ishida, T., Komatsu, H., Shudo, K., Kuwano, M., Ono, M., and Ueda, R. (2005). Antimyeloma effects of a novel synthetic retinoid Am80 (Tamibarotene) through inhibition of angiogenesis. Leukemia *19*, 901–909. https://doi.org/10.1038/sj.leu.2403754.
20. Ossowski, L., Russo, H., Gartner, M., and Wilson, E.L. (1987). Growth of a human carcinoma (HEp3) in nude mice: rapid and efficient metastasis. J. Cell. Physiol. *133*, 288–296. https://doi.org/10.1002/jcp.1041330212.
21. Bragado, P., Estrada, Y., Parikh, F., Krause, S., Capobianco, C., Farina, H.G., Schewe, D.M., and Aguirre-Ghiso, J.A. (2013). TGF-β2 dictates disseminated tumour cell fate in target organs through TGF-β-RIII and p38α/β signalling. Nat. Cell Biol. *15*, 1351–1361. https://doi.org/10.1038/ncb2861.
22. Ossowski, L., and Reich, E. (1983). Changes in malignant phenotype of a human carcinoma conditioned by growth environment. Cell *33*, 323–333. https://doi.org/10.1016/0092-8674(83)90414-2.
23. Aguirre Ghiso, J.A., Kovalski, K., and Ossowski, L. (1999). Tumor dormancy induced by downregulation of urokinase receptor in human carcinoma involves integrin and MAPK signaling. J. Cell Biol. *147*, 89–104. https://doi.org/10.1083/jcb.147.1.89.
24. Khalil, B.D., Sanchez, R., Rahman, T., Rodriguez-Tirado, C., Moritsch, S., Martinez, A.R., Miles, B., Farias, E., Mezei, M., Nobre, A.R., et al. (2022). An NR2F1-specific agonist suppresses metastasis by inducing cancer cell dormancy. J. Exp. Med. *219*, e20210836. https://doi.org/10.1084/jem.20210836.
25. Di Martino, J.S., Nobre, A.R., Mondal, C., Taha, I., Farias, E.F., Fertig, E.J., Naba, A., Aguirre-Ghiso, J.A., and Bravo-Cordero, J.J. (2022). A tumor-derived type III collagen-rich ECM niche regulates tumor cell dormancy. Nat. Can. (Ott.) *3*, 90–107. https://doi.org/10.1038/s43018-021-00291-9.
26. Aguirre-Ghiso, J.A., Ossowski, L., and Rosenbaum, S.K. (2004). Green fluorescent protein tagging of extracellular signal-regulated kinase and p38 pathways reveals novel dynamics of pathway activation during primary and metastatic growth. Cancer Res. *64*, 7336–7345. https://doi.org/10.1158/0008-5472.CAN-04-0113.
27. Aguirre-Ghiso, J.A., Estrada, Y., Liu, D., and Ossowski, L. (2003). ERK(-MAPK) activity as a determinant of tumor growth and dormancy; regulation by p38(SAPK). Cancer Res. *63*, 1684–1695.
28. Aslakson, C.J., and Miller, F.R. (1992). Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. Cancer Res. *52*, 1399–1405.
29. Stresemann, C., and Lyko, F. (2008). Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. Int. J. Cancer *123*, 8–13. https://doi.org/10.1002/ijc.23607.

30. Schenk, T., Stengel, S., and Zelent, A. (2014). Unlocking the potential of retinoic acid in anticancer therapy. Br. J. Cancer *111*, 2039–2045. https://doi.org/10.1038/bjc.2014.412.

31. Bigg, H.F., Mcleod, R., Waters, J.G., Cawston, T.E., and Clark, I.M. (2000). Mechanisms of induction of human tissue inhibitor of metalloproteinases-1 (TIMP-1) gene expression by all-transretinoic acid in combination with basic fibroblast growth factor. Eur. J. Biochem. *267*, 4150–4156. https://doi.org/10.1046/j.1432-1327.2000.01459.x.

32. Chen, H., Zhang, H., Lee, J., Liang, X., Wu, X., Zhu, T., Lo, P.-K., Zhang, X., and Sukumar, S. (2007). HOXA5 acts directly downstream of retinoic acid receptor β and contributes to retinoic acid–induced apoptosis and growth inhibition. Cancer Res. *67*, 8007–8013. https://doi.org/10.1158/0008-5472.can-07-1405.

33. Tang, X.-H., and Gudas, L.J. (2011). Retinoids, retinoic acid receptors, and cancer. Annu. Rev. Pathol. *6*, 345–364. https://doi.org/10.1146/annurev-pathol-011110-130303.

34. Nobre, A.R., Risson, E., Singh, D.K., Di Martino, J.S., Cheung, J.F., Wang, J., Johnson, J., Russnes, H.G., Bravo-Cordero, J.J., Birbrair, A., et al. (2021). Bone marrow NG2+/Nestin+ mesenchymal stem cells drive DTC dormancy via TGF-β2. Nat. Can. (Ott.) *2*, 327–339. https://doi.org/10.1038/s43018-021-00179-8.

35. Prunier, C., Baker, D., Ten Dijke, P., and Ritsma, L. (2019). TGF-Beta family signaling pathways in cellular dormancy. Trends Cancer *5*, 66–78. https://doi.org/10.1016/j.trecan.2018.10.010.

36. Salm, S.N., Burger, P.E., Coetzee, S., Goto, K., Moscatelli, D., and Wilson, E.L. (2005). TGF-{beta} maintains dormancy of prostatic stem cells in the proximal region of ducts. J. Cell Biol. *170*, 81–90. https://doi.org/10.1083/jcb.200412015.

37. Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., and Ma'Ayan, A. (2010). ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics *26*, 2438–2444. https://doi.org/10.1093/bioinformatics/btq466.

38. Akhtar-Zaidi, B., Cowper-Sal Lari, R., Corradin, O., Saiakhova, A., Bartels, C.F., Balasubramanian, D., Myeroff, L., Lutterbaugh, J., Jarrar, A., Kalady, M.F., et al. (2012). Epigenomic enhancer profiling defines a signature of colon cancer. Science *336*, 736–739. https://doi.org/10.1126/science.1217277.

39. Morrow, J.J., Bayles, I., Funnell, A.P.W., Miller, T.E., Saiakhova, A., Lizardo, M.M., Bartels, C.F., Kapteijn, M.Y., Hung, S., Mendoza, A., et al. (2018). Positively selected enhancer elements endow osteosarcoma cells with metastatic competence. Nat. Med. *24*, 176–185. https://doi.org/10.1038/nm.4475.

40. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936. https://doi.org/10.1073/pnas.1016071107.

41. Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320–334. https://doi.org/10.1016/j.cell.2013.03.036.

42. Massagué, J. (2012). TGFβ signalling in context. Nat. Rev. Mol. Cell Biol. *13*, 616–630. https://doi.org/10.1038/nrm3434.

43. Inman, G.J., Nicolás, F.J., and Hill, C.S. (2002). Nucleocytoplasmic shuttling of smads 2, 3, and 4 permits sensing of TGF-β receptor activity. Mol. Cell *10*, 283–294. https://doi.org/10.1016/s1097-2765(02)00585-3.

44. Massagué, J., Seoane, J., and Wotton, D. (2005). Smad transcription factors. Genes Dev. *19*, 2783–2810. https://doi.org/10.1101/gad.1350705.

45. Zhao, M., Mishra, L., and Deng, C.X. (2018). The role of TGF-beta/SMAD4 signaling in cancer. Int. J. Biol. Sci. *14*, 111–123. https://doi.org/10.7150/ijbs.23230.

46. Ossowski, L., and Reich, E. (1980). Experimental model for quantitative study of metastasis. Cancer Res. *40*, 2300–2309.

47. Bansal, N., Bosch, A., Leibovitch, B., Pereira, L., Cubedo, E., Yu, J., Pierzchalski, K., Jones, J.W., Fishel, M., Kane, M., et al. (2016). Blocking the PAH2 domain of Sin3A inhibits tumorigenesis and confers retinoid sensitivity in triple negative breast cancer. Oncotarget *7*, 43689–43702. https://doi.org/10.18632/oncotarget.9905.

48. Tobita, T., Takeshita, A., Kitamura, K., Ohnishi, K., Yanagi, M., Hiraoka, A., Karasuno, T., Takeuchi, M., Miyawaki, S., Ueda, R., et al. (1997). Treatment with a new synthetic retinoid, Am80, of acute promyelocytic leukemia relapsed from complete remission induced by all-trans retinoic acid. Blood *90*, 967–973.

49. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97. https://doi.org/10.1093/nar/gkw377.

50. Nagy, Á., Munkácsy, G., and Győrffy, B. (2021). Pancancer survival analysis of cancer hallmark genes. Sci. Rep. *11*, 6047. https://doi.org/10.1038/s41598-021-84787-5.

51. Bornstein, S., White, R., Malkoski, S., Oka, M., Han, G., Cleaver, T., Reh, D., Andersen, P., Gross, N., Olson, S., et al. (2009). Smad4 loss in mice causes spontaneous head and neck cancer with increased genomic instability and inflammation. J. Clin. Invest. *119*, 3408–3419. https://doi.org/10.1172/jci38854.

52. Ding, Z., Wu, C.J., Chu, G.C., Xiao, Y., Ho, D., Zhang, J., Perry, S.R., Labrot, E.S., Wu, X., Lis, R., et al. (2011). SMAD4-dependent barrier constrains prostate cancer growth and metastatic progression. Nature *470*, 269–273. https://doi.org/10.1038/nature09677.

53. Borgen, E., Rypdal, M.C., Sosa, M.S., Renolen, A., Schlichting, E., Lønning, P.E., Synnestvedt, M., Aguirre-Ghiso, J.A., and Naume, B. (2018). NR2F1 stratifies dormant disseminated tumor cells in breast cancer patients. Breast Cancer Res. *20*, 120. https://doi.org/10.1186/s13058-018-1049-0.

54. Andrews, S. (2019). https://github.com/s-andrews/FastQC.

55. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

56. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419. https://doi.org/10.1038/nmeth.4197.

57. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

58. Fontanals-Cirera, B., Hasson, D., Vardabasso, C., Di Micco, R., Agrawal, P., Chowdhury, A., Gantz, M., de Pablos-Aragoneses, A., Morgenstern, A., Wu, P., et al. (2017). Harnessing BET inhibitor sensitivity reveals AMIGO2 as a melanoma survival gene. Mol. Cell *68*, 731–744.e9. https://doi.org/10.1016/j.molcel.2017.11.004.

59. Hasson, D., Panchenko, T., Salimian, K.J., Salman, M.U., Sekulic, N., Alonso, A., Warburton, P.E., and Black, B.E. (2013). The octamer is the major form of CENP-A nucleosomes at human centromeres. Nat. Struct. Mol. Biol. *20*, 687–695. https://doi.org/10.1038/nsmb.2562.

60. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

61. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

62. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008).

Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

63. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.

64. Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319. https://doi.org/10.1016/j.cell.2013.03.035.

65. Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210. https://doi.org/10.1093/nar/30.1.207.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-Smad2	Cell Signaling Technology	Cat#5339S; RRID: AB_10626777
Rabbit monoclonal anti-Smad3	Cell Signaling Technology	Cat#9523P; RRID: AB_2193182
Rabbit monoclonal anti-Smad4	Cell Signaling Technology	Cat#38454; RRID: AB_2728776
Rabbit monoclonal anti-Smad4 (For IF)	Cell Signaling Technology	Cat#46535S; RRID: AB_2736998
Rabbit monoclonal anti-phospho-Smad2	Cell Signaling Technology	Cat#3108S; RRID: AB_490941
Rabbit monoclonal anti- phospho-Smad3	Cell Signaling Technology	Cat#9520P; RRID: AB_2193207
Rabbit monoclonal anti-Smad1	Cell Signaling Technology	Cat#9743S; RRID: AB_2107780
Rabbit monoclonal anti-Smad5	Cell Signaling Technology	Cat#12534S; RRID: AB_2797946
Rabbit monoclonal anti-phospho-Smad1/5	Cell Signaling Technology	Cat#9516S; RRID: AB_491015
Rabbit monoclonal anti-p27 Kip1	Cell Signaling Technology	Cat#3686S; RRID: AB_2077850
Mouse monoclonal anti-p27 Kip1 (For IF)	Cell Signaling Technology	Cat#3698S; RRID: AB_2077832
Mouse monoclonal anti-ERK1/2	BD Biosciences	Cat#610031; RRID: AB_397447
Mouse monoclonal anti-p38α	BD Biosciences	Cat#612168; RRID: AB_399540
Mouse monoclonal anti-Phospho-p38 MAPK	BD Biosciences	Cat#612288; RRID: AB_399605
Mouse monoclonal anti-phosphor-ERK1/2	Santa Cruz Biotechnology	Cat#sc7383; RRID: AB_627545
Rat monoclonal anti-vimentin	R & D Systems	Cat#MAB2105; RRID: AB_22141653
Rabbit monoclonal anti-Nr2f1	Abcam	Cat#ab181137; RRID: AB_2890250
Rabbit polyclonal anti-Lamin B1	Abcam	Cat#ab16048; RRID: AB_10107828
Mouse monoclonal anti-H3K27ac	EpiCypher	Cat#13–0045; RRID: AB_2923489
Mouse monoclonal anti-beta-tubulin	Sigma-Aldrich	Cat#T7816; RRID: AB_261770
Chemicals, peptides, and recombinant proteins		
Dulbecco's Modified Eagle Medium (DMEM)	Fisher Scientific	Cat#10-013-CV
Heat-inactivated fetal bovine serum (FBS)	Peak Serum	Cat#PS-FB1
Penicillin and Streptomycin	Fisher Scientific	Cat#30-002-CI
RPMI-1640	Fisher Scientific	Cat#SH30027FS
Collagenase-1A	Sigma Aldrich	Cat#C9891
5-Azacytidine	Sigma Aldrich	Cat#A2385
Charcoal-stripped FBS	Thermo Fisher	Cat#12676029
all-trans retinoic acid	Sigma Aldrich	Cat#R2625
AM80	Tocris	Cat#3507
Lipofectamine RNAiMax	Invitrogen	Cat#13778075
4X Laemmli sample buffer	Bio-Rad	Cat#1610747
NE-PER Nuclear and Cytoplasmic extraction reagent	Thermo Fisher	Cat#78833
M-MuLV reverse transcriptase enzyme	New England Biolabs	Cat#M0253S
PowerUP SYBR Green master mix	Applied Biosystems	Cat#A25741
Antiigen retrieval buffer	Invitrogen	Cat#00-4955-58
Normal goat serum	Gibco	Cat#PCN5000
ProLong Gold Antifade Mounting medium with DAPI	Invitrogen	Cat#P36931
Hematoxylin	Fisher	Cat#353032
Eosin	Fisher	Cat#284532
Organo/Limonene Mount	Sigma	Cat#O8015

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GIPZ lentiviral shRNA transduction starter kit	Horizon	Cat#RHS5086-EG4089
KAPA HiFi DNA polymerase	Roche	Cat#07958838001
AMPure XP beads	NEB	Cat#A63880
Critical commercial assays		
PCR Mycoplasma test kit	PromoCell	Cat#PK-CA91-1096
RNeasy mini kit	Qiagen	Cat#74104
Pierce BCA protein assay kit	Thermo Scientific	Cat#23227
Amersham ECL Western Blot Detection	GE Healthcare	Cat#RPN2106
KAPAHiFi HotStart HyperPrep kit	Kapa Biosystems	Cat#KK2601
QIAquick PCR purification kit	Qiagen	Cat#28106
Deposited data		
Raw and analyzed data	This paper	GEO: GSE181837; GSE181838; GSE181839
Raw and analyzed data	This paper; Khalil et al.[24]	GEO: GSE172115
HNSCC patient data	Nagy et al.[50]	https://kmplot.com/analysis/
Experimental models: Cell lines		
HNSCC cell line: T-HEp3	Ossowski et al.[20]	RRID: CVCL_JB77
HNSCC cell line: D-HEp3	Ossowski et al.[20]	RRID: CVCL_JB77
Breast cancer cell line: 4T1	ATCC	RRID: CVCL_0125
Experimental models: Organisms/strains		
Mouse: NU/J, strain#002019	The Jackson Laboratory	RRID: IMSR_JAX:002019
Egg: Specific Pathogen Free, premium plus, incubated	Charles River	Cat#10100331
Oligonucleotides		
NR2F1 siRNA#1:	Thermo Fisher	Cat#4390824
NR2F1 siRNA#2:	Thermo Fisher	Cat#4392420
NR2F1 siRNA#3:	Millipore Sigma	Cat#NM_005654 (SASI_Hs01_00095428)
NR2F1 siRNA#4:	Millipore Sigma	Cat#NM_005654 (SASI_Hs01_00095429)
SAMD4 siRNA: targeting sequence: GCAAUUGAAAGUUUGGUAA	Horizon	Cat#J-003902-09-0005
SMAD4 sh1: clone ID: V3LHS_359402. Sequence: AACTCTGTACAAAGACCGC	Horizon	Cat#RHS4430-200305036
SMAD4 sh2: clone ID: V3LHS_408444. Sequence: TAACTTTGAGAACTTCTCT	Horizon	Cat#RHS4430-200301270
SMAd4 sh3: clone ID: V3LHS_359404. Sequence: TCCAGGTGATACAACTCGT	Horizon	Cat#RHS4430-200302670
Scr sh: clone ID: E20200805A. Sequence: N/A	Horizon	Cat#RH4346
Q-PCR primers	IDT	See supplementary table
Software and algorithms		
ImageJ	Version 1.8.0/1.53a	https://ImageJ.nih.gov/ij/index.html; RRID:SCR_003070
CaseViewer	Version 2.4	https://www.3dhistech.com/solutions/caseviewer/; RRID:SCR_017654
fastQC	Version 0.11.8. Andrews, S.[54]	https://github.com/s-andrews/FastQC; RRID:SCR_014583
STAR	Version 2.7.5b. Dobin et al.[55]	https://github.com/alexdobin/STAR; RRID:SCR_004463
Trim Galore	Version 0.6.6	https://github.com/FelixKrueger/TrimGalore; RRID:SCR_011847

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Salmon	Version 1.11	https://github.com/COMBINE-lab/salmon/releases; RRID:SCR_017036
DEseq2	Version 1.34.0. Love et al.[57]	https://github.com/mikelove/DESeq2; RRID:SCR_015687
Bowtie2	Version 2.2.8. Langmead et al.[60]	https://github.com/BenLangmead/bowtie; RRID:SCR_005476
PICARD	Version 2.2.4	https://github.com/broadinstitute/picard/releases; RRID:SCR_006525
Samtools	Version 1.11. Li et al.[61]	http://www.htslib.org/; RRID:SCR_002105
MACS2	Version 2.1.0. Zhang et al.[62]	https://pypi.org/project/MACS2/; RRID:SCR_013291
deepTools	Version3.2.1. Ramirez et al.[63]	https://deeptools.readthedocs.io/en/develop/; RRID:SCR_016366
Ranking Ordering of Super Enhancers algorithm (ROSE)	Whyte et al.[64]; Loven et al.[41]	http://younglab.wi.mit.edu/super_enhancer_code.html; RRID:SCR_017390
RStudio	Version 1.1.46	https://rstudio.com/; RRID:SCR_000432
R	Version 4.1.0	https://www.r-project.org/; RRID:SCR_001905
GE ImageQuant LAS 4010	GE Healthcare	https://www.gehealthcare.com/
GraphPad Prism	Version 9	https://www.graphpad.com/; RRID:SCR_002798

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Julio A. Aguirre-Ghiso (julio.aguirre-ghiso@einsteinmed.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- RNA-seq and ChIP-seq data have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resource table. This paper analyzes existing, publicly available data of HNSCC patients, and the source information is listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines and primary culture

Tumorigenic (T-HEp3) cells were derived from a lymph node metastasis in head and neck squamous cell carcinoma (HNSCC) patient and maintained as patient-derived xenograft on chicken chorioallantoic membrane (CAM) as described below and previously.[22,23] Dormant (D-HEp3) cells were obtained by *in vitro* passaging of T-HEp3 cells for more than 40 generations. Both cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) with 10% heat-inactivated fetal bovine serum (FBS) and 100U penicillin/0.1 mg/mL streptomycin. 4T1 cells were maintained in RPMI-1640 medium containing 10% FBS and 100U penicillin/0.1 mg/mL streptomycin. Cells were incubated and grown at 5% CO2 and 37°C. Cells were routinely tested for mycoplasma using PCR Mycoplasma test kit.

In vivo chicken chorioallantoic membrane (CAM) model

Use of the chicken embryo was approved by the Institutional Animal Care and Use Committee (IACUC) of the Albert Einstein College of medicine. Briefly, 150 × 10^3 T-HEp3, 500 x 10^3 D-HEp3 and 250 x 10^3 4T1 cells were inoculated on CAM and allowed to grow. After

7 days, tumors were harvested and digested with collagenase-1A (Sigma-Aldrich; C9891) for 30 min at 37°C. Tumor cells (recognized by their very large size compared with chicken cells) were counted using a hematocytometer.[22,23] To check the effect of gene-specific knockdown by CAM assay, cells were transfected with gene-specific siRNAs in cell culture for 48 h and then inoculated to chicken CAM.

In vivo mouse model

For mouse experiments, 8-10-week-old NU/J female mice (strain# 002019) purchased from the Jackson Laboratories were used. Use of female nude mice and experimental procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of the Icahn School of Medicine at Mount Sinai, and the Albert Einstein College of Medicine.

METHOD DETAILS

Chicken chorioallantoic membrane (CAM) assay

CAM assay was performed in fertilized chicken eggs.[22,23] Cells in tissue culture were collected using PBS + 2mM EDTA (for T-HEp3 and D-HEp3) or trypsin (for 4T1) and concentration of viable cells were determined by trypan blue staining. Required number of cells were resuspended in 1X PBS (with calcium and magnesium). CAM of the chicken eggs (day 10 of development) at the top were dropped by puncturing the top of the shell and side air bag of the shell, and sucking air out from the side air bag. T-HEp3 (150×10^3), D-HEp3 (500×10^3) and 4T1 (250×10^3) cells resuspended in 50 μL was inoculated on CAM using fine tip pipette and incubated at 37°C in a humid chamber for 7 days. After 7 days (day 17 of chicken embryo development), tumors were harvested from the eggs, minced and digested with collagenase-1A (0.15%) + BSA (2.5%) in 1X PBS (with calcium and magnesium) for 30 min at 37°C. Collagenase digestion was terminated by adding equal volume of culture medium. Tumor cells were collected by centrifugation and resuspended in culture medium. The total number of cells (tumor cells recognized by their large size compared with chicken cells) per tumor was estimated by counting cells using a hematocytometer. *Note:* To check the effect of gene-specific knockdown by CAM assay, cells were transfected with gene-specific siRNAs in cell culture for 48 h, then inoculated to chicken CAM and tumor was collected after 7 days for RNA expression analysis by qPCR.

AZA and atRA reprogramming

T-HEp3 cells (P1) growing in a dish were treated with 5 nM 5-Azacytidine (AZA) in DMEM containing charcoal-stripped 10% FBS and Pen/Strep. After 24 h, AZA was replenished and allowed to grow for another 24 h. Next day, AZA-containing medium was removed, and cells were treated with 1 μM all-trans retinoic acid in serum-free medium with Pen/Strep and allowed to grow for another 48 h.

RNA sequencing and analysis

Total RNA from T-HEp3 cells reprogrammed with AZA+atRA was extracted using RNeasy mini kit following manufacturer's instructions. RNA samples QC, library preparation and RNA sequencing were performed at GENEWIZ, LLC. (South Plainfield, NJ, USA). Concentration of RNA samples were estimated by Qubit 2.0 Fluorometer (Life Technologies) and RNA integrity (RIN) was checked by Agilent TapeStation 4200 (Agilent Technologies). Poly-A enriched RNA was used to prepare RNA-seq libraries using the NEBNext Ultra RNA Library Prep Kit from Illumina following the manufacturer's protocol (New England Biolabs). The samples were sequenced in a 2 x 150 bp paired-end configuration using the Illumina HiSeq 4000 platform. Library preparation and RNA sequencing (Illumina HiSeq 2500) for D-HEp3 and T-HEp3 cell lines were performed at the Genomics Core Facility at the Icahn School of Medicine at Mount Sinai. Raw sequence data generated from Illumina HiSeq 4000 were converted to FASTQ files and demultiplexed using Illumina bcl2fastq 2.17 software. RNA-seq raw data were processed and analyzed by the BiNGS bioinformatics core at Icahn School of Medicine at Mount Sinai. Quality control was performed using FastQC (v0.11.8).[54] Trim Galore (v0.6.6) was used to trim the adapter sequence with a quality threshold of 20. Reads were aligned to the human reference genome (GRCh38.p13), and alignments were performed using STAR aligner (v2.7.5b).[55] Gene level read counts were obtained by Salmon (v1.11).[56] Assigned reads were then normalized and DEGs were identified using the R package DEseq2 (v1.34.0).[57] Differentially expressed genes were called using an adjusted p value <0.05 (using the Benjamini and Hochberg procedure). Principal component analysis was generated using regularized log-transformed reads with the DEseq2 package. Volcano plot was generated with the ggplot2 v3.3.5 package using DEseq2 results statistics. Heatmaps were generated with the pheatmap v1.0.12 package using DEseq2 median-ratio normalized counts. Pathway enrichment analysis was performed using Enrichr.[49]

ChIP sequencing and analysis

H3K27ac crosslinked ChIP was performed in D-HEp3, T-HEp3, T-HEp3-Control and T-HEp3-AZA+atRA (AA) cells following the method previously described[58] with several modifications. About 10 million cells per sample were cross-linked with 1% formaldehyde for 10 min at room temperature. Cross-linked cells were quenched with 0.125M glycine for 5 min at room temperature, and cells were collected by centrifuging at 400g for 3 min at 4°C. Cells were washed with ice-cold 1X PBS and resuspended in 500 μL cell lysis buffer (10 mM Tris pH 8.0, 10 mM NaCl, 0.2% NP-40, 100 nM PMSF, protease inhibitors), and incubated on ice for 15 min. Next, lysed cells were centrifuged at 400g for 5 min, and pellet was resuspended in 500 μL ice-cold nuclear lysis buffer (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS, 100 nM PMSF, protease inhibitors) followed by incubation for 10 min on ice. Next, cells were sonicated for 15–20

cycles, 30 s ON, 30 s OFF, at low intensity in a Bioruptor sonicator (Diagenode). Sonicated samples were centrifuged at 13,000g for 10 min at 4°C, and the supernatant with chromatin fraction was diluted four times with IP dilution buffer (20 mM Tris pH 8.0, 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.01% SDS, 100 mM PMSF, protease inhibitors). Drosophila chromatin was added as a spike-in to the diluted chromatin followed by pre-clearing with pre-conjugated IgG for 2h at 4°C gently rocking. 50 μL pre-cleared chromatin was saved for input control. Pre-conjugated H3K27ac antibody (100 μL slurry and 10 μg antibody per 10 million cells) were added to the pre-cleared chromatin and gently rotated for 16-18 h at 4°C. Next, beads were washed once with ice-cold IP wash buffer I (20 mM Tris pH 8.0, 2 mM EDTA, 50 mM NaCl, 1% Triton X-100, 0.1% SDS, 100 nM PMSF, protease inhibitors), twice with ice-cold high salt buffer (20 mM Tris pH 8.0, 2 mM EDTA, 500 mM NaCl, 1% Triton X-100, 0.01% SDS, 100 nM PMSF, protease inhibitors), once with ice-cold IP wash buffer II (10 mM Tris pH 8.0, 1 mM EDTA, 0.25 LiCl, 1% NP-40, 1% deoxycholic acid, 100 nM PMSF, protease inhibitors), followed by two washes with ice-cold TE buffer (5 mM Tris pH 7.4, 1 mM EDTA) twice. DNA was eluted twice from beads in 100 μL elution buffer (1% SDS, 100 mM $NaHCO_3$) at 65°C for 30 min shaking. To de-crosslink the DNA from ChIP fraction, 12 μL of 5 M NaCl, 2 μL of RNase A (10 mg/mL) were added, and for input, 130 μL TE buffer, 12 μL 5 M NaCl, 20 μL 10% SDS and 2 μL of RNase A were added followed by overnight incubation at 65°C. Next day, 4 μL of Proteinase K (20 mg/mL) were added and samples were incubated for 2h at 42°C. DNA was purified using the QIAquick PCR purification kit (Qiagen) following manufacturer's instructions. Libraries were generated and barcoded for multiplexed sequencing following Illumina's recommendations with minor modifications. Briefly, 2–8 ng input and ChIP DNA was end-repaired and A-tailed with deoxyadenosine. Illumina Truseq adapters were ligated, and libraries were PCR-amplified using KAPA HiFi DNA polymerase and HyperPrep kit with no more than 12–15 cycles. Libraries were size-selected to exclude polynucleosomes using AMpure beads. All steps in library preparation were carried out using NEB enzymes.[59] Libraries were sequenced on Illumina's NextSeq500 at 75 bp single-end reads. Sequencing reads were trimmed with Trim Galore v0.6.6 with default parameters, and aligned to the human reference genome hg38 using Bowtiev2.2.8[60] with parameters -end-to-end, -sensitive, -X 2000 and read quality was assessed using fastQC. Duplicate reads were removed with PICARD v2.2.4 (Broad Institute). Binary alignment maps (BAM) files were generated with samtools v1.11[61] and were used in downstream analysis. MACS2 v2.1.0[62] was used to called significant peaks. Peaks within ENCODE blacklisted regions were removed. Coverage tracks were generated from BAM files using deepTools v3.2.1[63] bam coverage with parameters– normalize using RPKM–bin size 10. For genomic annotation promoters (−1 kb to +1 kb) relative to the TSS were defined according to the human hg38 genome version. Heatmaps of genomic regions were generated with deepTools v3.2.1. The command compute matrix was used to calculate scores at genomic regions and generate a matrix file to use with plot heatmap, to generate plots.

ChIP-seq differential enrichment analysis

For D-HEp3 vs. T-HEp3 and T-HEp3 Control vs. T-HEp3 AA H3K27ac ChIP-seq, the BAM files from each comparison were merged in a single BAM and significant peaks were called using MACS2 narrow Peak v2.1.0 to generate a bed file with common set of regions. The BAM file of all the common regions was used to call enhancers using the Ranking Ordering of Super Enhancers algorithm (ROSE)[41,64] and the BAM files of the T-HEp3/D-HEp3 and T-HEp3 Control/T-Hep-3 AA were used to map their specific reads at the ROSE called enhancers. Using ROSE normalized and scaled counts, differential enhancers were called if the log2 fold change was greater than 1.5.

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus[65] and are accessible through GEO series accession number GSE181839 and GEO accession number GSE172115.

RNA interference

Gene-specific siRNAs (listed in key resources table) were transfected twice at the interval of 24 h to HEp3 cells at a final concentration of 50 nM, using Lipofectamine RNAiMax reagent (Invitrogen).

Western blot

Cells were lysed in RIPA buffer (50 mM Tris pH 7.4, 1 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.25% sodium deoxycholate, o.1% SDS, 1% NP40) and protein concentrations were calculated using Pierce BCA protein assay kit and a standard BSA curve. Samples were boiled in 4X laemmli sample buffer for 10 minutes at 95°C. To prepare nuclear protein extracts for immunoblotting, NE-PER nuclear and cytoplasmic extraction reagents was used and manufacturer's protocol was followed. 10-12% SDS–PAGE gels were run in running buffer (25 mM Tris, 190 mM glycine, 0.1% SDS) and transferred to PVDF membranes in transfer buffer (25 mM Tris, 190 mM glycine, 20% methanol). Membranes were blocked in 5% milk or BSA in TBST (Tris-buffered saline with 0.1% Tween-20) buffer. Membranes were incubated with primary antibodies overnight at 4°C. Membranes were washed in TBST buffer and incubated with HRP-conjugated secondary antibodies at room temperature for 1 hour. Western blots were developed using Amersham ECL Western Blot Detection and images were taken using GE ImageQuant LAS 4010. The antibodies used are listed in key resources table.

Quantitative PCR

RNA was extracted using RNeasy mini kit and cDNA was synthesized using M-MuLV reverse transcriptase enzyme (NEB) following the manufacturer's instructions. Real-time quantitative PCR was performed using PowerUP SYBR Green Master Mix in QuantStudio 3 (Applied Biosystems, A28567). Primers used for amplification are listed in Table S7.

Xenograft mouse model studies

Eight-week old female BALB/c nu/nu mice were injected subcutaneously with 0.5 x 10^6 T-HEp3-GFP cells. Mice were inspected regularly for arising tumors and when tumors reached ~300 mm^3, mice were treated with vehicle or one cycle of neoadjuvant AZA+atRA or AZA+AM80. 1 mg/kg/day of AZA and atRA and 0.3 mg/kg/day of AM80 doses were used for each treatment cycle (2 days of AZA + 3 days of atRA or AM80 + 2 days rest). After the neoadjuvant cycle, mice were injected with anesthetics ketamine (80–120 mg) and xylazine (5 mg), and tumors (~600 mm^3) were removed by surgery. Forty-eight hours post-surgery, mice were treated for four adjuvant cycles with the vehicle, AZA+atRA or AZA+AM80. At the end of the fourth cycle, mice were sacrificed by euthanization, and lungs were collected.

In the second mouse experiment, 2.5 x 10^5 T-HEp3 (scr, Smad4 sh1 and sh2 stable cell line; listed in key resources table) cells treated with DMSO (control) or AZA+atRA (reprogrammed) in-vitro, were injected to 8-week old female BALB/c nu/nu mice through tail vein. Starting next day, animals were treated with AZA (1 mg/kg/day) for 2 days followed by atRA (1 mg/kg/day) for next 2 days with next three days of no treatment. This treatment was given for 2 weeks by i.p. injection. At the end of the experiment, mice were sacrificed by euthanization, lungs were collected, fixed with 4% formaldehyde and processed for staining.

Immunofluorescence

Paraffin-embedded tissue sections were incubated in xylene (10 minutes twice) followed by graded ethanol rehydration (3 minutes each), and finally washed with water for 5 min twice. Antigen retrieval for mouse lung tissues was performed in 10 mM antigen retrieval buffer (pH 6) for 40 minutes using a steamer. Tissues were permeabilized in 0.3% TritonX-100 + PBS for 10 minutes. Cultured cells were fixed in 4% formaldehyde on ice for 10 minutes. Tissue sections and cultured cells were blocked with 3% bovine serum albumin (BSA, Fisher Bioreagents) and 5% normal goat serum in PBS for 1 hour at room temperature. Primary antibodies (listed in key resources table) in blocking buffer were incubated overnight at 4°C followed by washing with PBS three times and incubation with Alexa Fluor conjugated secondary antibodies (Invitrogen, 1:1000 dilution) at room temperature for 1-2 hour in the dark. Tissues were washed with PBS three times. Slides were mounted with ProLong Gold Antifade reagent with DAPI. Images were obtained using Leica Software on a Leica SPE confocal microscope and analyzed using ImageJ software.

Hematoxylin and eosin staining and image analysis

Paraffin-embedded tissue sections were incubated in xylene (10 minutes twice) followed by graded ethanol rehydration (3 minutes each), and finally washed with water for 5 min twice. Next, tissue sections were stained with hematoxylin for 30 sec and washed immediately with running water for 5 min, followed by eosin staining for 1 min. After a short dip in water, tissue sections were dehydrated with graded ethanol (70% for 2 min, 95% for 2 min, 100% for 2 min twice, and xylene for 5 min twice. Tissue sections were mounted in Organo/Limonene mounting medium and dry overnight. H & E stained tissue sections were scanned using Fast Scanner at Imaging Facility at Albert Einstein College of Medicine. Scanned images were analysed for metastasis by Caseviewer v2.4.

Survival analysis

The regression-free survival analyses were performed using RNA-seq data of KM plotter (https://kmplot.com/analysis/) in HNSCC patients. Patient were stratified into two groups as per to various quantile expression of the mean expression of multiple genes examined. The difference between cohorts were characterized by hazard ratio with 95% confidence intervals and the statistical significance (log rank) was computed using Cox-Mantel test.

QUANTIFICATION AND STATISTICAL ANALYSIS

All *in vitro* experiments were repeated at least three times unless indicated otherwise. For CAM tumor growth analysis, a minimum of 3 tumors were analyzed per group/experiment. For mouse experiments, a minimum of 15 mice per group were used for tumor growth, while a minimum of 5 mice per group were used for immunostaining analysis. Statistical analysis was performed on Prism software using unpaired t-test, Mann- Whitney test, and 2-way ANOVA with Holm Sidak's multiple comparison test. A p-value $\leq$ 0.05 was considered significant.