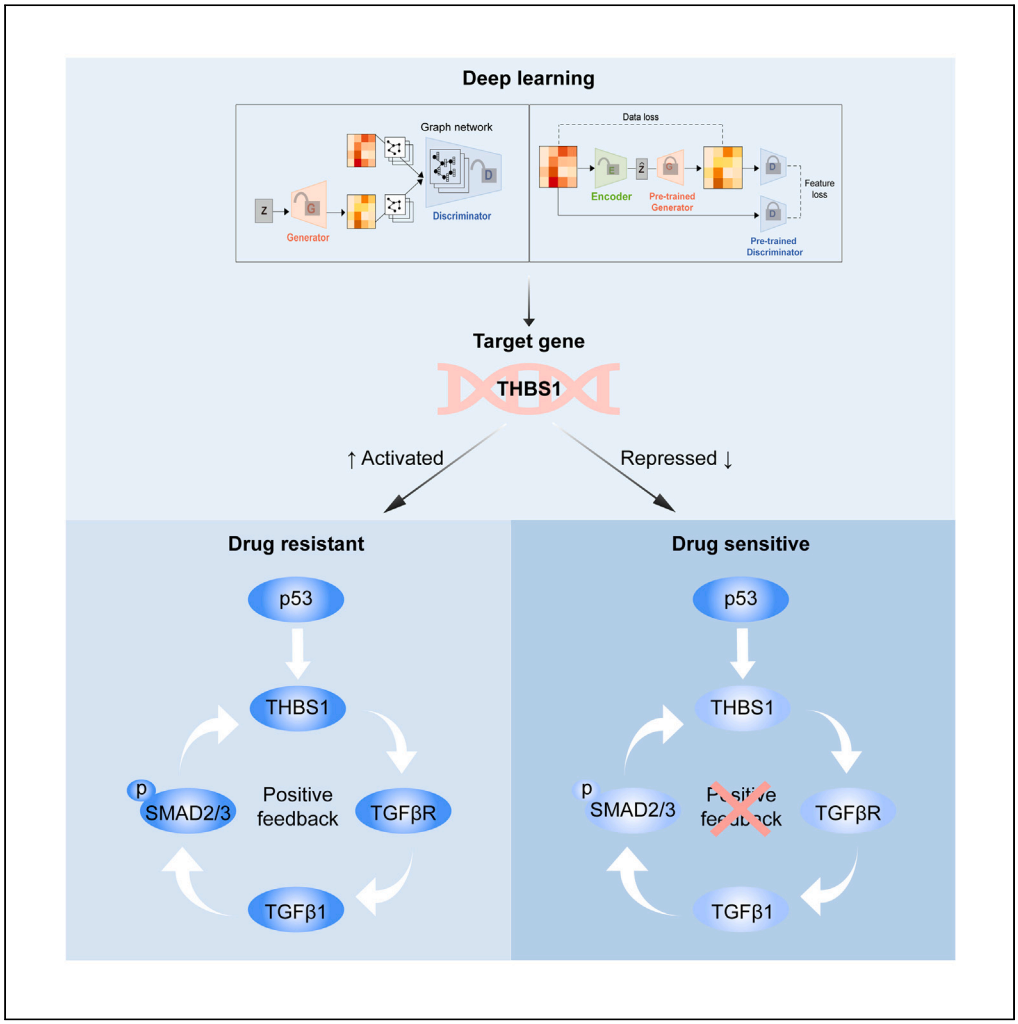
iScience

CellPress
OPEN ACCESS

Article

Deep learning untangles the resistance mechanism of p53 reactivator in lung cancer cells



Soo Min Lee, Younghyun Han, Kwang-Hyun Cho

ckh@kaist.ac.kr

Highlights

We present AnoDAN, a deep learning model to unravel the drug resistance mechanism

AnoDAN identifies a target gene by integrating gene and pathway-level information

A loop between the TGF-β pathway and THBS1 causes p53 reactivator resistance

Inhibiting THBS1 overcomes p53 reactivator resistance in lung cancer cells

Lee et al., iScience 26, 108377
December 15, 2023
https://doi.org/10.1016/j.isci.2023.108377

iScience

CellPress
OPEN ACCESS

Article

Deep learning untangles the resistance mechanism of p53 reactivator in lung cancer cells

Soo Min Lee,[1,2] Younghyun Han,[1,2] and Kwang-Hyun Cho[1,3,*]

SUMMARY

Tumor suppressor p53 plays a pivotal role in suppressing cancer, so various drugs has been suggested to upregulate its function. However, drug resistance is still the biggest hurdle to be overcome. To address this, we developed a deep learning model called AnoDAN (anomalous gene detection using generative adversarial networks and graph neural networks for overcoming drug resistance) that unravels the hidden resistance mechanisms and identifies a combinatorial target to overcome the resistance. Our findings reveal that the TGF-β signaling pathway, alongside the p53 signaling pathway, mediates the resistance, with THBS1 serving as a core regulatory target in both pathways. Experimental validation in lung cancer cells confirms the effects of THBS1 on responsiveness to a p53 reactivator. We further discovered the positive feedback loop between THBS1 and the TGF-β pathway as the main source of resistance. This study enhances our understanding of p53 regulation and offers insights into overcoming drug resistance.

INTRODUCTION

Cancer is a major contributor to global mortality, and its incidence is projected to increase over the next 50 years due to the expansion of aging.[1] Lung cancer is the second most common type of cancer, yet it has the highest mortality rate among all cancer types, according to the World Health Organization and the World Cancer Research Fund International. Lung cancer comprises two main types, which are non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC can be further sub-classified into adenocarcinoma, squamous cell carcinoma, and large cell carcinoma based on histology, as reported by the Lung Cancer Foundation of America. Among the diverse types, lung adenocarcinoma (LUAD) is the most prevalent, and there have been many efforts to develop effective therapies for LUAD patients.[2]

LUAD is a complex disease caused by multifactorial factors. However, one of the prominent factors is the abnormality of the tumor suppressor genes. Among them, TP53 is a critical tumor suppressor gene that has an important role in preventing cancer development. Studies based on The Cancer Genome Atlas (TCGA) datasets have indicated that TP53 is the most commonly mutated gene in LUAD.[3,4] The wild-type TP53 gene remains inactive due to its interaction with the negative regulator MDM2 until it encounters any signal or stress. Upon stimulation, TP53 binds to DNA sequences through its DNA binding domain and promotes cell-cycle arrest, DNA repair, and apoptosis to suppress the progression of tumors. Unfortunately, TP53 mutations frequently occur in the DNA-binding domain, which then leads to loss of function and tumor progression.[5] Thus, restoring the wild-type function of the mutant TP53 is of utmost importance.

Several drugs have been proposed to restore p53 function, including those that target MDM2 or MDM4 negative regulators or those that directly focus on mutant p53 reactivation. Mutant p53 reactivators include Chetomin, CP-31398, MIRA-1, NSC87511, PRIMA-1, and PRIMA-1MET, but most of them have failed to receive FDA approval due to solubility issues and toxicity to normal cells.[6] MDM2 inhibitors include Nutlins, AMG232, CGM097, and HDM201, and MDM4 inhibitors are NSC207895 and 17-AAG.[7] However, these inhibitors are only effective when the MDM2 or MDM4 gene is amplified and when p53 is wild type. One of the compounds that act as a direct reactivator of mutant p53 is PRIMA-1MET, also known as APR-246 or Eprenetapopt. It is the only compound that is under advanced clinical trials and is a more active and permeable form of PRIMA-1 with the addition of a methyl group. It is first converted to methylene quinuclidinone (MQ), and it reacts with thiol groups and modifies the thiol groups in mutant p53.[8] As MQ gains its ability to bind to cysteine residues in the DNA-binding domain of the mutant p53 protein, it reactivates the p53 wild-type conformation and restores its tumor suppressor functions.[5] With this ability, APR-246 has a great possibility for efficacy in mutant p53 patients.

Although APR-246 is a promising drug for treating cases with mutant p53, drug resistance presents a significant challenge that requires further investigation. To overcome the resistance of APR-246 and enhance its efficacy, numerous studies have attempted to identify potential combinatorial targets or drugs. For instance, in lung cancer, studies have shown that cisplatin and olaparib have synergistic effects with APR-246.[9–11] APR-246 has also been found to help overcome resistance to cisplatin and doxorubicin in ovarian cancer,[12] while SLC1A11 has been

[1]Laboratory for Systems Biology and Bio-inspired Engineering, Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea
[2]These authors contributed equally
[3]Lead contact
*Correspondence: ckh@kaist.ac.kr
https://doi.org/10.1016/j.isci.2023.108377

identified as a significant contributor to APR-246 resistance.[13] Knockdown of MRP1 has also been found to increase sensitivity to APR-246 in mutant p53 cell lines.[14] However, uncovering the mechanisms of drug resistance, rather than just identifying target genes, is a challenging yet necessary task.

The aim of this study is to comprehend the mechanism underlying APR-246 resistance in LUAD and to identify a target that can overcome it. Although analyzing differentially expressed genes (DEGs) between sensitive and resistant groups has been previously conducted, it has limitations since drug resistance results from complex molecular interactions at the pathway-level.[15] Likewise, while deep learning has been used to find therapeutic biomarkers, the underlying mechanism is elusive.[16] Some studies have implemented explainable deep learning models, but there is more to explore than just finding out which features are influential to make a prediction. To address these challenges, we propose a computational framework called AnoDAN (anomalous gene detection using generative adversarial networks (GAN) and graph neural networks (GNN) for overcoming drug resistance). AnoDAN combines GAN and GNN to identify genes that can overcome drug resistance and to unravel the underlying mechanisms by incorporating pathway information. It can also accurately reproduce the distribution of gene expressions in the sensitive cell lines and identify anomalous genes in the resistant cell lines. Furthermore, AnoDAN can incorporate datasets with graphical structures using the biological pathway information of the genes to discover hidden mechanisms as well as target genes for overcoming the resistance. Here, we identified the p53 pathway and TGF-β pathway as central pathways and THBS1 as a target gene to overcome resistance to APR-246 in LUAD. We also revealed resistance mechanisms and further validated drug response after regulating THBS1 through *in vitro* experiments. Together, our study presents a combinatorial target in p53 reactivator-resistant lung cancer cells and provides insights into the mechanisms underlying drug resistance.

RESULTS

Model training results

We employed the AnoDAN framework to discover a combinatorial gene that can overcome drug resistance. The study design is illustrated in Figure 1. Our investigation focused on APR-246, a reactivator of mutant p53, and we categorized the cell lines into two groups: APR-246 sensitive and resistant, based on their inhibitory concentration (IC50) values. AnoDAN consists of three essential components, namely a generator, discriminator, and encoder (Figure 2A). To capture the distinctive characteristics of sensitive cell lines, AnoDAN was trained using gene expression data from these cells. The anomaly score was computed using the trained AnoDAN by calculating difference between real and generated gene expression data to analyze pathway-level mechanisms and identify the target gene (Figure 2B). Subsequently, *in vitro* experiments were conducted to validate the effects of the target gene on drug response and elucidate the underlying mechanisms. Detailed information about the architecture of the AnoDAN can be found in the STAR methods.

The model training was first confirmed using the loss functions, and the results indicated that the generator, discriminator, and encoder were sufficiently trained as evidenced by converging losses (Figure 2C). The Uniform Manifold Approximation and Projection (UMAP) plot, which is a common method for reducing high-dimensional data into a two-dimensional representation,[17] was also used to verify model training. The degree of overlap between the real and generated data points is used to assess whether a GAN model is sufficiently trained. In this study, the real sensitive and generated data points at the final epoch overlapped substantially, signifying that the GAN underwent effective training (Figure 2D). The UMAPs at the initial and intermediate steps of training showed the gradual progress in learning process of the model, as shown in Figure S1A. However, in contrast to these results, the real resistant and generated data points did not exhibit the same level of overlap observed with the real sensitive points. This implies that the model primarily learned the distributions of sensitive cell line data only (Figure S1B). Furthermore, it is reasonable to speculate that the resistant cell lines might display higher average sample-level anomaly scores, considering that the model was trained using APR-246 sensitive cell line data. As anticipated, our results showed that the resistant cell lines indeed display significantly higher anomaly scores compared to the sensitive cell lines (Figure S1C). Therefore, we validated that the model is sufficiently trained to capture the characteristics of the sensitive cell lines.

AnoDAN provides three distinct advantages over other methods. Firstly, it incorporates more biological knowledge as it integrates pathway information during training by utilizing GNN. Secondly, it identifies the key gene that causes major differences between sensitive and resistant cell lines. Thirdly, it enables the simultaneous identification of target genes and the interpretation of its mechanism with respect to pathways. Of note, DEG analysis only compares gene expression values of two distinct groups without providing any explanation for the results, whereas AnoDAN incorporates pathway information. For this reason, the resulting list of genes was also completely different as seen in Table 1. Furthermore, while other deep learning approaches focus on identifying therapeutic biomarkers, they often fail to elucidate the mechanism behind the discovery of a biomarker.[16] Thus, compared to popular tools like DEG analysis and other deep learning methods, AnoDAN is a distinct framework for identifying combinatorial targets and their related mechanisms.

p53 and TGF-β signaling pathways are primarily responsible for resistance

Following model training, anomaly scores can be determined at the pathway-level, sample-level, and gene-level through the anomality scoring process. The resulting pathway-level anomaly score revealed that the top-scored pathways are the p53 signaling pathway, the TGF-β signaling pathway, and the SCLC (Figure 3A). The scores ranged from 20 to 40 when the highest score was set to 100. From the score distribution plot, top-scored pathways were classified as those with scores exceeding 60 (Figure S2A). Accordingly, the scores for the top pathways were 100, 75.4, and 66.2, respectively. The p53 signaling pathway was identified as a top pathway, which is consistent with the fact that the APR-246 is a mutant p53 reactivator. Additionally, previous studies suggested that the TGF-β signaling pathway could potentially

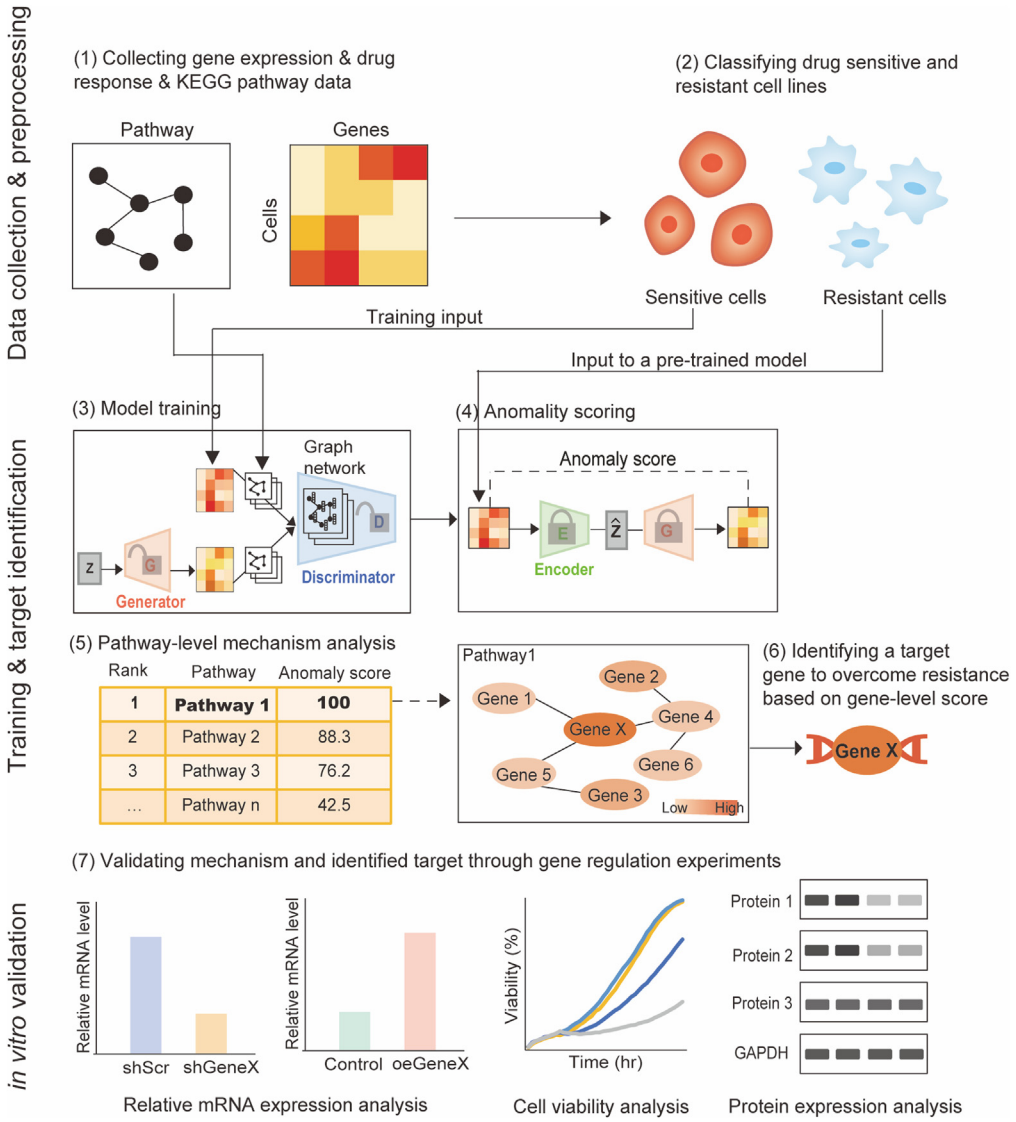


Figure 1. Overview of the study design

Our study comprises three principal stages: data collection and preprocessing, model training and target identification, and *in vitro* validation.
(1) We first collect gene expression, drug response, and pathway data.
(2) We then classify cell lines into drug sensitive and resistant cell lines.
(3) Subsequently, our model is trained utilizing gene expression and pathway data of drug sensitive cells.
(4) Anomaly scores are computed using the pre-trained model.
(5) Based on the obtained pathway-level anomaly scores, we analyze pathways that are predominantly implicated in conferring resistance.
(6) We then explore gene-level anomaly scores in the top-scored pathways to identify a target gene for overcoming resistance.
(7) Finally, we experimentally validate the resistance mechanism and identified target through *in vitro* experiments.

play a significant role in the development of resistance to targeted therapy.[18] Furthermore, p53 and TGF-β signaling pathways are interlinked, as mutant p53 may impede TGF-β's tumor-suppressing role by failing to activate its target genes.[19] Therefore, the pathway-level score results and related studies underscore the importance of the p53 and TGF-β signaling pathways in mutant p53 reactivator resistance.

The reliability of AnoDAN is validated through three distinct analyses

To assess the reliability of AnoDAN and the resulting findings, the model was evaluated through variation in application, diversity in dataset, and graph-based evaluation. First, we subjected the pipeline to validation using a different drug, Trametinib, an MEK inhibitor. This drug is known to target MEK1 and MEK2 and is associated with the MAPK signaling pathway. The model was sufficiently trained as presented in Figures S3A and S3B. The pathway-level score results demonstrated that the TGF-β signaling pathway, the Notch signaling pathway, and the MAPK signaling

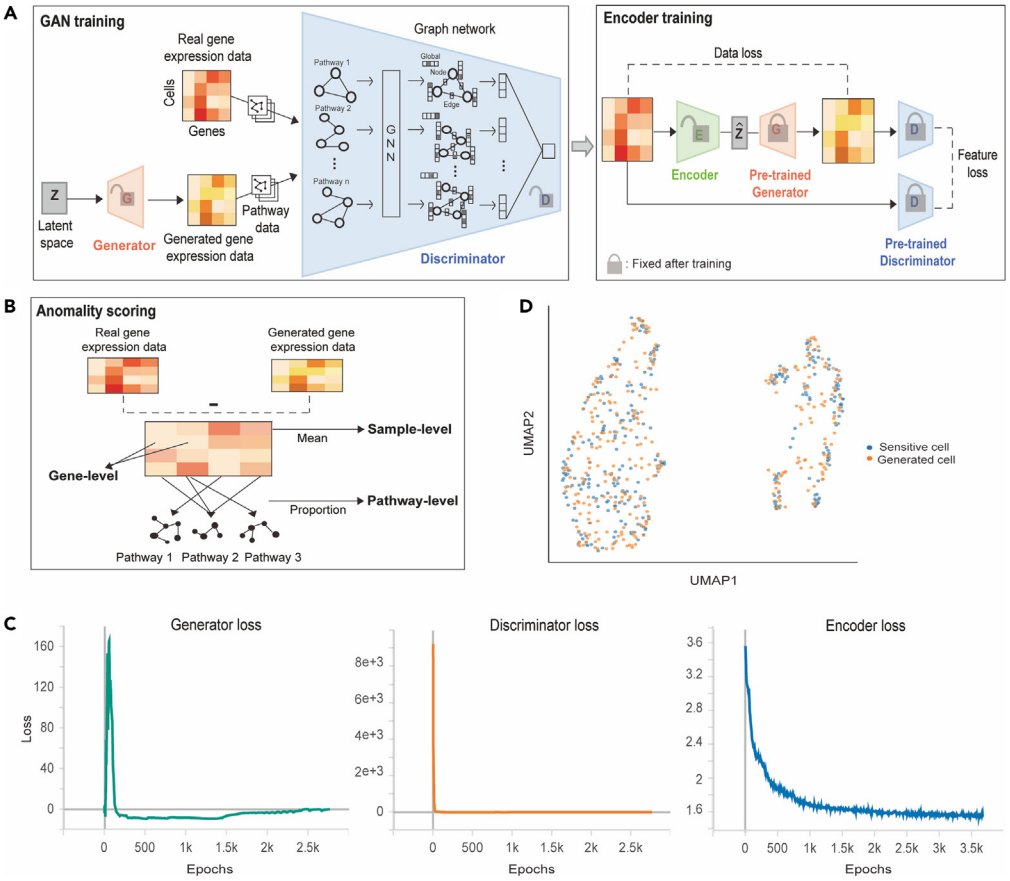


Figure 2. Architecture and training results of AnoDAN

(A) Schematic representation of the AnoDAN architecture. AnoDAN training consists of GAN and Encoder training. Gene expression and pathway data of sensitive cell lines are mapped and employed in the GAN training process. Graph neural networks are implemented in the discriminator. The pre-trained generator and discriminator are subsequently utilized to train the encoder, which also leverages sensitive cell line data.

(B) Detailed visualization of anomality scoring. Anomaly scores can be computed by comparing real gene expression with generated gene expression data. Scores are obtained at three levels: sample, gene, and pathway-level.

(C) Loss curves for the generator, discriminator, and encoder.

(D) UMAP plot of the GAN model. The resulting data points, comprising real sensitive data (blue) and generated data (orange), are reduced to two dimensions.

pathway are mainly responsible for resistance (Figure S3C). Some studies proved that TGF-β and MAPK signaling pathways are concurrently related to drug resistance with the HAT1 gene.[20] TGF-β signaling-mediated resistance was also studied with BRAFi/MEKi, and the pathway was speculated to be a key factor in trametinib resistance.[18] Co-targeting the MAPK/ERK and Notch signaling pathways were also suggested as a strategy to improve current cancer therapies.[21] In addition to the related pathways, several genes, including MEK1, MEK2, ERK1, ERK2, BRAF, and KRAS, that are related to trametinib and its resistance were further investigated. The expression of KRAS gene has been known to be related to MEK inhibitor resistance.[22] BRAF is an upstream gene of MEK and ERK1/2 genes are downstream genes of MEK in the MAPK signaling pathway. As a result, anomaly scores of the six genes were higher in the resistant cell lines compared to the sensitive cell lines (Figure S3D). This result demonstrates that AnoDAN, which is trained with the trametinib dataset, accurately classifies the genes attributed to trametinib resistance. Based on these results, we validated that AnoDAN is also applicable to other drugs.

It is also essential to evaluate the robustness of our findings by utilizing various independent datasets. The GDSC database that we used in this study contains the most comprehensive APR-246 data, and the Cancer Therapeutics Response Portal v2 (CTRPv2) database stands as the second-largest dataset in this context (see STAR methods). Therefore, we applied the pretrained AnoDAN model to CTRPv2 dataset to compute the anomaly score of the APR-246 resistant cell lines. The model identified four pathways, the p53 signaling pathway, the TGF-β signaling pathway, the SCLC, and the NSCLC, as the highest-scored pathways, which is consistent with our initial results (Figure S3E). By observing two different datasets with the consistent results, we concluded that our results are robust irrespective of datasets used.

In the model training process, pathway information was incorporated to impart more biological information to the model. Therefore, we further investigated whether the model trained on APR-246 sufficiently utilizes the connection links among genes in each pathway by

Table 1. Top-ranked differentially expressed genes (DEGs) between APR-246 resistant and sensitive cell lines

Gene Symbol	logFC	AveExpr	t	adj.P.Value	B
CSF2RA	1.23740	3.85350	2.82225	0.00787	−3.24631
PCK2	0.76977	5.95040	2.27766	0.02907	−3.77123
LRP5	0.75472	4.24637	2.37357	0.02334	−3.68311
EIF4EBP1	0.54974	8.70917	2.35514	0.02435	−3.70020
MLST8	0.47007	6.04464	2.69971	0.01069	−3.36915
VANGL1	0.43680	6.13314	2.52377	0.01639	−3.54107
MAPKAPK5	0.32306	7.53149	2.18176	0.03605	−3.85719
RASGRF2	0.16946	3.25953	3.06103	0.00426	−3.00087
ARHGEF6	0.15539	3.02520	2.17317	0.03674	−3.86477
CD3E	0.15162	2.99387	3.07349	0.00412	−2.98787
IL19	0.13995	3.06820	2.71453	0.01030	−3.35442
PSTPIP1	0.11781	3.31235	2.22915	0.03243	−3.81499
CYP7A1	−0.07239	2.87883	−2.03951	0.04914	−3.98035
LCP2	−0.10999	2.97049	−2.15206	0.03849	−3.88334
MTMR8	−0.11121	3.09484	−2.16546	0.03737	−3.87157
PIP4K2B	−0.26733	5.19828	−2.07416	0.04561	−3.95084
STAT5B	−0.26778	4.20323	−2.15645	0.03812	−3.87948
MTMR4	−0.28316	5.63926	−2.39577	0.02216	−3.66241
HSPA8	−0.31152	4.55383	−2.61582	0.01312	−3.45182
SACM1L	−0.35700	5.77553	−2.70353	0.01058	−3.36535
MAPK8	−0.37040	4.33737	−2.16354	0.03753	−3.87325
PPP2CB	−0.37850	9.75489	−2.11046	0.04216	−3.91958
CTNNB1	−0.39876	8.16499	−2.08867	0.04420	−3.93839
INPP5A	−0.41030	5.20685	−2.27919	0.02897	−3.76984
RNASEL	−0.42793	3.65431	−3.11374	0.00371	−2.94578
PRKAR1A	−0.43274	8.28995	−2.54195	0.01569	−3.52357
CASP7	−0.44555	5.04226	−2.07257	0.04577	−3.95220
PPM1D	−0.45548	5.28652	−2.58377	0.01419	−3.48308
ACAA1	−0.48158	6.25164	−2.71216	0.01036	−3.35678
SMAD4	−0.50934	5.90347	−2.74884	0.00946	−3.32019
EGLN1	−0.56473	6.44071	−2.63359	0.01257	−3.43442
LPIN3	−0.59207	4.71826	−2.18743	0.03560	−3.85216
MGST3	−0.64006	10.35586	−2.36405	0.02386	−3.69195
SMAD2	−0.75956	6.73912	−2.34836	0.02474	−3.70647
HBEGF	−1.01496	4.77739	−2.34577	0.02488	−3.70887
BMP2	−1.82161	4.74875	−3.65146	0.00086	−2.37093

comparing the results before and after shuffling the edges. As a result, top pathways obtained after training with the original structure were different after edge shuffling. Also, the absolute anomaly scores for the p53 and the TGF-β signaling pathway decreased after shuffling five times (Figures S3F and S3G). This result suggests that the top two pathways were obtainable with our model because pathway information was essentially used in the training to identify pathways that are highly responsible for resistance. The results obtained through the application to trametinib, the use of CTRPv2 dataset, and the shuffling of edges increased the reliability of our method and findings.

THBS1 is identified as a combinatorial target to overcome APR-246 resistance

The top two signaling pathways, p53 and TGF-β signaling pathways, were further investigated by examining the anomaly scores of each gene in the pathways. Gene-level anomaly scores were measured in the LUAD context, and the scores in each pathway were represented in a KEGG pathway graph. In the p53 signaling pathway, CDKN2A, IGFBP3, SERPINB5, and THBS1 genes had the highest scores (Figure 3B). In the

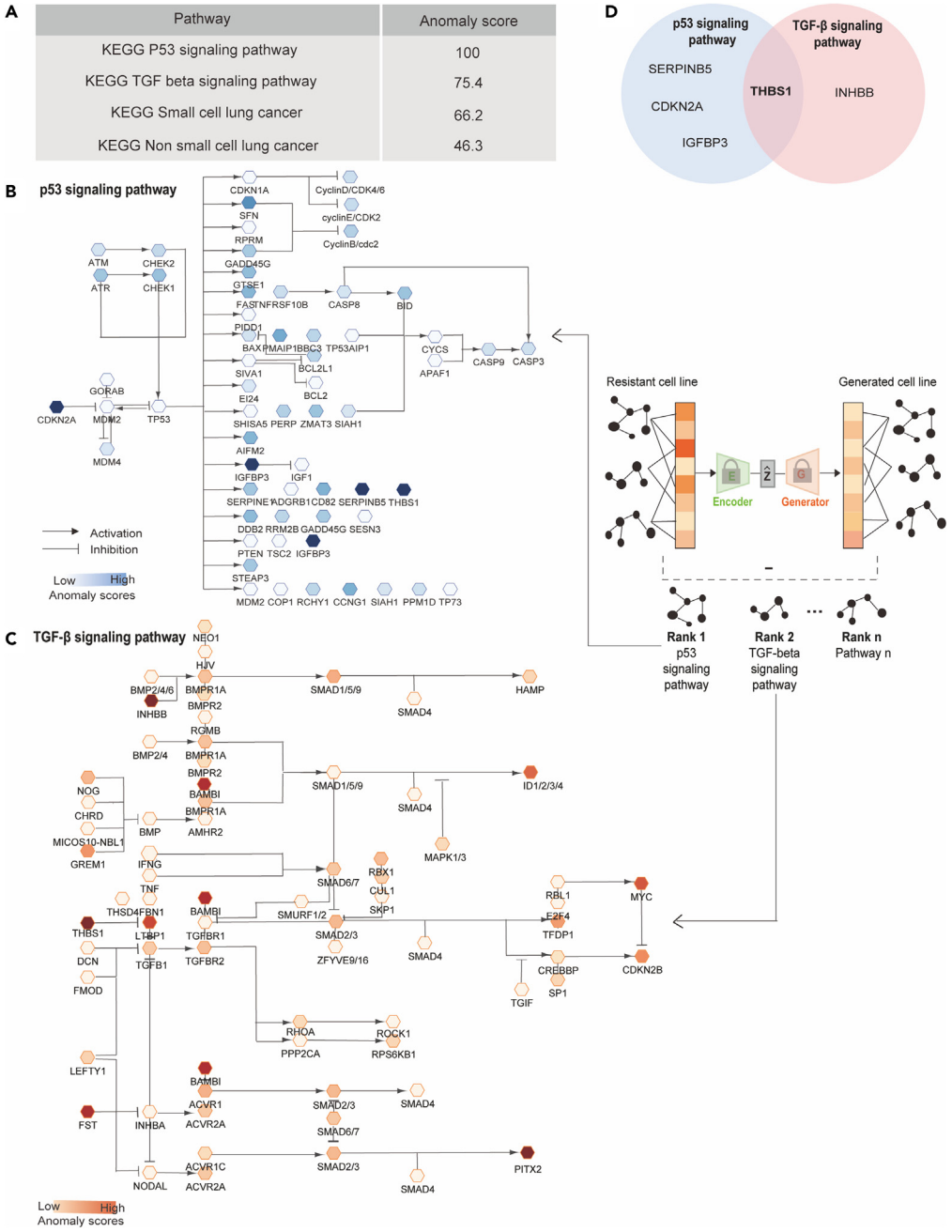


Figure 3. Identification of pathways and genes responsible for resistance

(A) Results of pathway-level anomaly scores.

(B) Network visualization of the p53 signaling pathway.

(C) Network visualization of the TGF-β signaling pathway. These two pathways ranked first and second, respectively, in the anomality scoring procedure. Each node represents a gene, with the depth of color denoting its anomaly score. Nodes with higher scores exhibit darker colors.

(D) Venn diagram depicting top-scored genes in the p53 and TGF-β signaling pathways. The overlapping gene is situated at the center of the diagram.

TGF-β pathway, INHBB and THBS1 had the highest scores, as represented by the darkest color (Figure 3C). The top-scored genes are represented in Figure 3D. Gene-level scores were enriched between zero and one, so scores bigger than 1.5 was considered to be the top score (Figure S2B). For the top-scored gene, their relations with APR-246 or p53-related drugs were additionally scrutinized. A study suggested that PRIMA-1 treatment with retained CDKN2A shows significant growth inhibition in human glioblastoma and that resistance to PRIMA-1 occurs through the loss of CDKN2A mRNA and protein expression.[23] For IGFBP-3, it was found to play a critical role in responsiveness to DNA

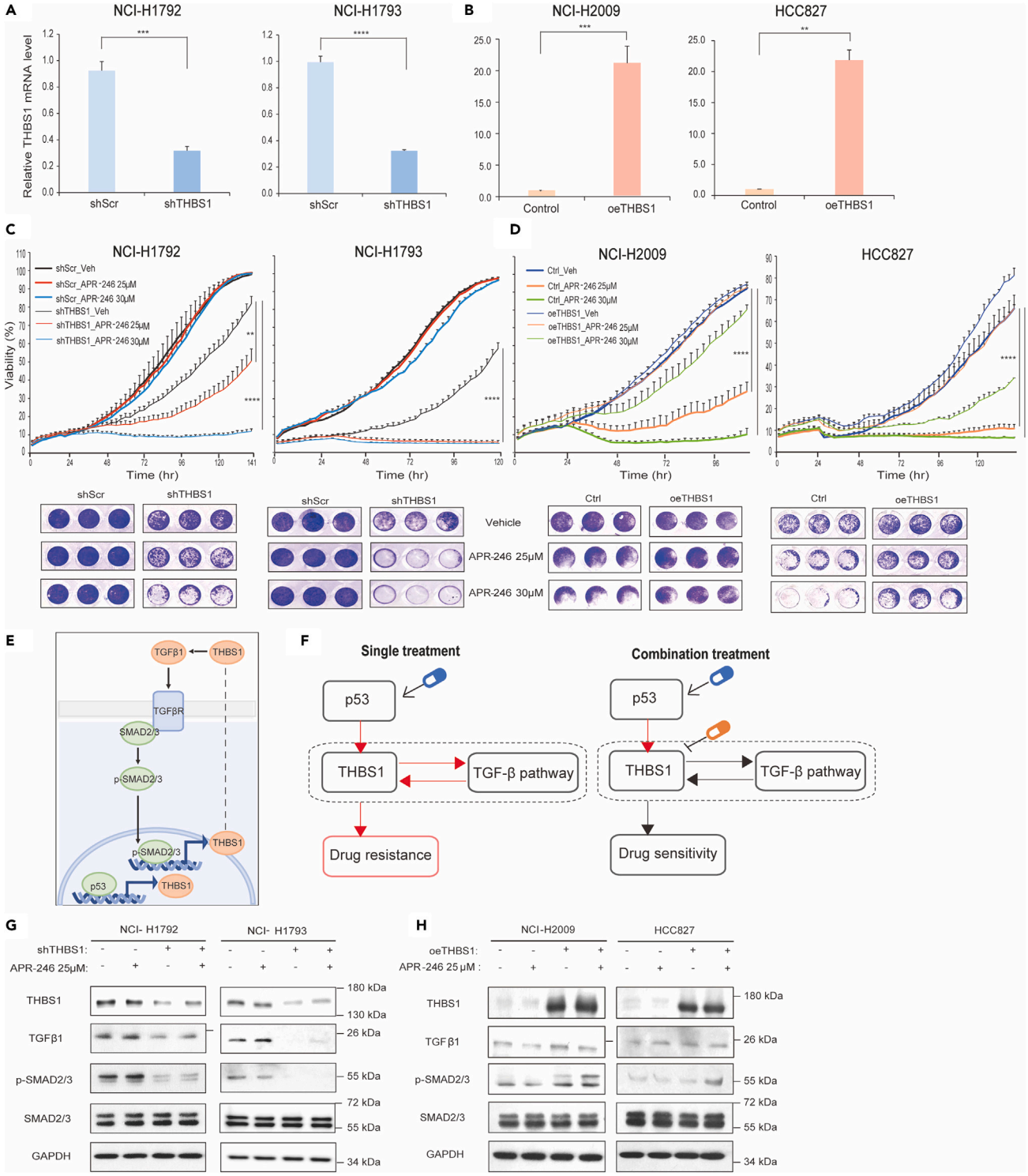


Figure 4. Experimental validation of the resistance mechanism and identified target

(A) THBS1 Knockdown in resistant cell lines.

(B) THBS1 overexpression in sensitive cell lines. Data are presented as mean ± SD. *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$ (Student's *t* test). Scramble shRNA, shScr; THBS1 shRNA, shTHBS1; THBS1 overexpression, oeTHBS1; Vehicle, Veh.

(C) Cell viability analysis in resistant cells with reduced THBS1 levels.

Figure 4. *Continued*

(D) Cell viability analysis in sensitive cells with elevated THBS1 levels. Data are presented as mean ± SD of three replicates. *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$ (Student's *t* test). Cells were seeded in a 96-well plate and treated with DMSO, APR-246 25 μM, or 30 μM after 24 h. Colonies were stained with crystal violet as depicted in the bottom.

(E) Schematic representation of the interplay between the TGF-β signaling pathway, THBS1, and p53. The dotted line denotes autocrine effects of THBS1.

(F) Abstracted representation of the resistance mechanism under monotherapy and combination treatment. Red lines or boxes indicate activation.

(G) Western blot analysis in resistant cells after knockdown.

(H) Western blot analysis in sensitive cells after overexpression. Cells were subjected to the indicated gene regulation or treatments, and the specified proteins were detected.

damaging therapy.[24] In addition, APR-246 was found to have a strong synergy with DNA damaging drugs such as cisplatin and doxorubicin.[13] Therefore, regulating IGFBP-3 may have a combinatorial effect with the APR-246 treatment. SERPINB5, THBS1, and INHBB were possible combinatorial targets with APR-246. Out of the three genes, THBS1 is the only one involved in both the p53 pathway and the TGF-β pathway, so it was expected to have a wider influence on cells and a stronger effect to overcome resistance.

THBS1, which is also known as TSP1 and Thrombospondin 1, is primarily involved in processes such as angiogenesis, metastasis, apoptosis, and cell-to-cell and cell-to-matrix interactions.[25] It often plays dual roles, and there is research on glioblastoma suggesting that THBS1 silencing inhibits tumor cell invasion and growth.[26] Similarly, a study conducted in lung osteosarcoma demonstrated that THBS1 silencing suppresses cell wound healing, migration, and invasion and inhibits pulmonary metastasis.[27] Based on this evidence, THBS1 was identified as a combinatorial target for APR-246.

Gene expression data obtained from GDSC showed significant increase in THBS1 expression in resistant cell lines with a p value of 1.75e-10 (Figure S2C), and our real-time qPCR analysis demonstrated that the mRNA level of THBS1 was higher in resistant cell lines than to sensitive cell lines (Figure S2D). Thus, we postulated that the downregulation of THBS1 would enhance the APR-246 drug response in resistant cell lines.

Regulating THBS1 level leads to different sensitivities to APR-246

Experimental validation of a target gene was performed using shRNA and overexpression vectors for THBS1 knockdown and overexpression, respectively. Prior to viral production, the drug responses of NCI-H1792, NCI-H1793, NCI-H2009, and HCC827 cell lines were experimentally validated. Cell lines were selected based on *Z* score normalized IC50 values derived from GDSC data and the presence of TP53 mutations. Results from cell growth assays indicated that NCI-H1792 and NCI-H1793 exhibited increased cell viability while NCI-H2009 and HCC827 exhibited statistically significant cell death upon treatment with APR-246 (Figures S4A and S4B). Treatment was administered at concentrations of 10, 15, 20, 25, and 30 μM with a control (vehicle), and drug response was dose-dependent in sensitive cell lines. Crystal violet staining of APR-246 treated cells corroborated cell growth curve data and showed a clear distinction between sensitive and resistant cell lines (Figure S4C). Consequently, NCI-H1792 and NCI-H1793 were designated as resistant cell lines while NCI-H2009 and HCC827 were designated as sensitive cell lines throughout the paper.

Knockdown of THBS1 was performed using shRNA, as shown in Figure 4A, with primers listed in Table 2. Cell growth curves were compared between shScr and shTHBS1. At an APR-246 concentration of 30 μM, cell viability was maintained at 10% in both NCI-H1793 and NCI-H1792. Statistically significant decreases in cell viability were observed in knockdown cells compared to the control at a concentration of 25 μM with a p value of 0.002 in NCI-H1792 and 7.61e-10 in NCI-H1793 (Figure 4C). Crystal violet staining of living cells revealed a marked difference in THBS1 knockdown cells. The same experiments were conducted using different THBS1 shRNA sequence to avoid the potential off-target effects. For those treated with different shRNA sequence, the drug response also significantly increased in NCI-H1792 and NCI-H1793 (Figures S5A and S5B). Therefore, these results indicate that reduction of THBS1 levels in resistant cells can overcome resistance to APR-246 and increase drug sensitivity. Additionally, the THBS1 gene was overexpressed using an overexpression vector in NCI-H2009 and HCC827, which are the APR-246 sensitive cell lines (Figure 4B). Cell viability assays showed that the cell line became resistant to APR-246 after overexpression of THBS1 in both cell lines (Figure 4D). This indicates that THBS1 is a critical mediator of resistance to APR-246.

Additionally, THBS1 knockdown was performed in sensitive cell lines such as NCI-H2009 and HCC827. Drug responses were enhanced at lower APR-246 concentrations, as shown in Figures S5C and S5D. Drug toxicity is a common issue in cancer treatment, but combinatorial treatment involving THBS1 silencing and APR-246 administration at a lower concentration in sensitive cell lines may mitigate this problem.

Resistance can be overcome by inhibiting the positive feedback loop between THBS1 and TGF-β pathway

The activation of signaling pathways has a strong influence on diseases or infections,[28] and positive feedbacks in a pathway are critical in elucidating the underlying mechanisms.[29] Also, the genes involved in positive feedbacks often act as master switches to drive biological events.[30] In this study, the p53 signaling pathway and the TGF-β signaling pathways are related in such a way that p53 activates THBS1, which in turn activates the TGF-β pathway through TGFβ1 activation. Previous studies have demonstrated that THBS1 activates the TGF-β pathway via phosphorylation of SMAD2/3.[31,32] Once the TGF-β pathway is activated, it transcriptionally regulates THBS1[26] as well as represented in Figure 4E. This positive feedback between the THBS1 and TGF-β pathway underlies the resistance mechanism observed in this study.

The TGF-β signaling pathway is widely associated with drug resistance and has been shown to inhibit tumor cell progression in early stages but promote it in advanced stages.[18] In various cancer types, including NSCLC, the TGF-β signaling pathway is strongly associated with resistance to targeted therapy. Overexpression of TGFβRII has been shown to induce drug resistance through activation of

Table 2. real-time qPCR primer sequences

Target gene	Forward primer sequence (5′-3′)	Reverse primer sequence (5′-3′)
GAPDH	TGATGACATCAAGAAGGTGGTGAAG	TCCTTGGAGGCCATGTGGGCCAT
THBS1	GGGGCGTCAATGACAATTTCCAG	TCACCACGTTGTTGTCAAGGGT
TGFβ1	TACCTGAACCCGTGTTGCTCTC	GTTGCTGAGGTATCGCCAGGAA

the TGF-β signaling pathway.[33,34] However, resistance mechanisms vary depending on cancer type and drug and are not yet fully understood.

Here, we hypothesized that the positive feedback loop between THBS1 and the TGF-β signaling pathway results in adverse effects that outweigh the therapeutic benefits of APR-246 under monotherapy (Figure 4F). This is attributable to the pre-existing activation of THBS1 via TGF-β, as represented in Figure 4E. Consequently, downregulation of THBS1 expression attenuates TGF-β pathway activation, thereby potentiating the efficacy of APR-246 (Figure 4F).

Consistent with our hypothesis, western blot analysis confirmed that concomitant THBS1 silencing and APR-246 treatment in resistant cell lines resulted in diminished expression of THBS1, TGFβ1, and phospho-SMAD2/3 proteins (Figure 4G). The identical results were observed with different shRNA sequence targeting THBS1 (Figure S5E). We also demonstrated that resistant cell lines exhibiting elevated THBS1 levels also displayed higher TGFβ1 levels relative to sensitive cell lines (Figure S6A). THBS1 knockdown cells exhibited reduced TGFβ1 mRNA levels, corroborating our previous results (Figure S6B). Subsequently, we derived pathway activity scores from Signaling Pathway Enrichment using Experimental Datasets (SPEED) provided by GDSC. A comparison between resistant and sensitive cell lines utilized in our study revealed a statistically significant increase in SPEED TGF-β signaling pathway activity in resistant cell lines with a p value of 0.0394 (Figure S6C). Collectively, these data substantiate our assertion that the TGF-β signaling pathway is hyperactivated in resistant cell lines.

Conversely, sensitive cell lines exhibiting lower THBS1 and TGFβ1 levels were treated with a TGF-β pathway activator for 24 h. As anticipated, this resulted in an elevation of TGFβ1 levels in both sensitive cell lines, NCI-H2009 and HCC827, and the THBS1 level also increased with the same tendency (Figures S6D and S6E). Western blot analysis performed on control, THBS1 overexpressing, APR-246 treated, and APR-246 untreated cells revealed that TGFβ1 and phospho-SMAD2/3 protein expressions increased in parallel with THBS1 expression (Figure 4H). These findings confirm that activation of the TGF-β signaling pathway induces THBS1 activation; thus, out experimental data validated the existence of a positive feedback loop between THBS1 and the TGF-β pathway. Therefore, inhibition of THBS1 overcomes resistance by disrupting this feedback mechanism.

DISCUSSION

LUAD is the most prevalent subtype of lung cancer, with TP53 mutation being the most common. Numerous TP53-related drugs have been developed, with some targeting wild-type p53 and others mutant p53, and most of them leading to upregulation of p53. Despite their potential, drug resistance poses a significant obstacle to their effectiveness. This study specifically focuses on APR-246, a mutant p53 reactivator that restores the wild-type conformation and tumor suppressor functions of p53. It is the only mutant p53 reactivator in the advanced stages of clinical trials.[35] Despite being used in various cancer types, APR-246 faces a drug resistance problem that hinders its effectiveness. Therefore, overcoming resistance to such p53 reactivators is the primary goal of this study.

To find a target gene that plays a significant role in resistance, we developed a computational framework called AnoDAN. In order to mimic biologically plausible gene expression patterns by machine learning models, we combined GAN and GNN. The GNN in the discriminator enabled the generator to incorporate more biologically relevant information into the generated samples. By integrating GNN into the GAN, we were able to identify which signaling pathways are key mediators of resistance, providing more insights into the biological mechanisms underlying drug resistance.

Consequently, we found that the p53 and the TGF-β signaling pathways are mostly responsible for APR-246 resistance with the highest pathway-level anomaly scores. The THBS1 gene was overlapped in both pathways, as its gene-level anomaly score was one of the top-ranked scores for the two pathways in common, therefore, we selected it as a combinatorial target with APR-246. From *in vitro* experiments, we validated that knockdown of THBS1 increases its responsiveness to APR-246 in resistant cell lines and that overexpression of THBS1 induces resistance in sensitive cell lines. The primary mechanism was the positive feedback loop between THBS1 and the TGF-β pathway, which ultimately mediates the sensitivity to APR-246.

AnoDAN is a versatile computational framework that can also be utilized to identify combinatorial targets for overcoming resistance to various drugs, since many targeted therapies confront drug resistance problems. For instance, when AnoDAN was trained on gene expression data from cell lines sensitive to trametinib (an MEK inhibitor), it identified the MAPK pathway, which is strongly associated with trametinib, as one of the top pathways (Figure S3C). This indicates that AnoDAN can be widely utilized for other drugs, particularly for those with limited information on effective combinatorial targets. Moreover, this framework, which involves comparing two groups with distinct characteristics and detecting anomalies, has previously been employed in different fields, such as traffic flow detection, retinal damage detection, network intrusion detection, and IoT big data anomaly detection.[36] However, aside from being applied to medical image data, this approach has not been applied in biological fields, particularly when using gene expression and pathway data. In addition, a variety of methodologies exist for

modeling of signaling pathways and their dynamics.[37] These pathways serve a pivotal role in elucidating biological phenomena such as cell state transitions across distinct cellular states.[38,39] Given the significance of signaling pathways in the context of drug resistance, our framework comprehensively addresses the pathway-related components of resistance mechanisms. Therefore, efforts to apply this approach in different contexts, such as the identification of combinatorial targets through the integration of gene expression and pathway data, hold substantial significance and offer opportunities for addressing numerous unsolved problems in the field of cancer research.

In summary, discovering a combinatorial target that enhances the efficacy of a drug is essential to prevent frequent cancer treatment failures. In this study, we suggest a combination of the p53 reactivator, e.g., APR-246, and THBS1 silencing as a therapeutic strategy in mutant p53 LUAD. We also propose a computational framework called AnoDAN to find an anomalous gene in resistant lung cancer cells. In addition, AnoDAN can be applied to different drugs, including other p53 upregulating drugs, and tissues to discover cancer therapy combinations and decipher underlying resistance mechanisms.

Limitations of the study

This study predominantly utilized gene expression data for the identification of a target gene to overcome resistance. However, it is important to acknowledge that certain target genes may exhibit no discernible differences in expression. Despite their unaltered expression, these targets may still hold substantial significance in overcoming drug resistance. Furthermore, some targets which do not have a direct causal link with resistance but still exert a significant influence on resistance may remain undetected by our method. Another limitation of our study is that identification of a target gene is limited to the genes involved in specific pathways utilized during training. To address these limitations, further studies should undertake more comprehensive analyses, incorporating diverse data types such as DNA, protein, and epigenetics data, and encompassing a broader spectrum of pathway information.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell culture
  - Reagents
- METHOD DETAILS
  - Data preparation
  - Architecture and training procedure of AnoDAN
  - Anomality scoring
  - Virus production for knockdown and overexpression experiments
  - Cell growth analysis and crystal violet assay
  - Total RNA extraction and quantitative real-time PCR (qRT-PCR) analysis
  - Western blot analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108377.

ACKNOWLEDGMENTS

We would like to thank all SBiE (Systems Biology and Bio-inspired Engineering) members for their critical comments on the research. We also thank Chae Young Hwang for her supportive comments on the experiments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government, the Ministry of Science and ICT (2023R1A2C3002619), the grant of Electronics and Telecommunications Research Institute (ETRI) funded by the Korean government [23ZS1100, Core Technology Research for Self-Improving Integrated Artificial Intelligence System], and the grant of the Korea Dementia Research Project through the Korea Dementia Research Center (KDRC), funded by the Ministry of Health & Welfare and Ministry of Science and ICT, Republic of Korea (grant number: HU21C0060).

AUTHOR CONTRIBUTIONS

S.M.L., Y.H., and K.-H.C. conceptualized the research. S.M.L. performed data analysis and experiments. Y.H. conducted data analysis and provided analytical guidance. K.-H.C. designed the study and supervised the research. S.M.L., Y.H., and K.-H.C. wrote and reviewed the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 1, 2023
Revised: September 12, 2023
Accepted: October 27, 2023
Published: November 1, 2023

REFERENCES

1. Soerjomataram, I., and Bray, F. (2021). Planning for tomorrow: global cancer incidence and the role of prevention 2020-2070. Nat. Rev. Clin. Oncol. *18*, 663–672. https://doi.org/10.1038/s41571-021-00514-z.
2. Ferone, G., Lee, M.C., Sage, J., and Berns, A. (2020). Cells of origin of lung cancers: lessons from mouse studies. Genes Dev. *34*, 1017–1032. https://doi.org/10.1101/gad.338228.120.
3. Cancer Genome Atlas Research Network (2014). Comprehensive molecular profiling of lung adenocarcinoma. Nature *511*, 543–550. https://doi.org/10.1038/nature13385.
4. Chang, Y.S., Tu, S.J., Chen, Y.C., Liu, T.Y., Lee, Y.T., Yen, J.C., Fang, H.Y., and Chang, J.G. (2021). Mutation profile of non-small cell lung cancer revealed by next generation sequencing. Respir. Res. *22*, 3. https://doi.org/10.1186/s12931-020-01608-5.
5. Berke, T.P., Slight, S.H., and Hyder, S.M. (2022). Role of Reactivating Mutant p53 Protein in Suppressing Growth and Metastasis of Triple-Negative Breast Cancer. OncoTargets Ther. *15*, 23–30. https://doi.org/10.2147/OTT.S342292.
6. Hassin, O., and Oren, M. (2023). Drugging p53 in cancer: one protein, many targets. Nat. Rev. Drug Discov. *22*, 127–144. https://doi.org/10.1038/s41573-022-00571-8.
7. Zhu, Y., Wang, H., and Thuraisamy, A. (2019). MDM2/P53 Inhibitors as Sensitizing Agents for Cancer Chemotherapy. In Protein Kinase Inhibitors as Sensitizing Agents for Chemotherapy (Elsevier), pp. 243–266.
8. Lambert, J.M.R., Gorzov, P., Veprintsev, D.B., Söderqvist, M., Segerbäck, D., Bergman, J., Fersht, A.R., Hainaut, P., Wiman, K.G., and Bykov, V.J.N. (2009). PRIMA-1 reactivates mutant p53 by covalent binding to the core domain. Cancer Cell *15*, 376–388. https://doi.org/10.1016/j.ccr.2009.03.003.
9. Perdrix, A., Najem, A., Saussez, S., Awada, A., Journe, F., Ghanem, G., and Krayem, M. (2017). PRIMA-1 and PRIMA-1(Met) (APR-246): From Mutant/Wild Type p53 Reactivation to Unexpected Mechanisms Underlying Their Potent Anti-Tumor Effect in Combinatorial Therapies. Cancers *9*, 172. https://doi.org/10.3390/cancers9120172.
10. Deben, C., Lardon, F., Wouters, A., Op de Beeck, K., Van den Bossche, J., Jacobs, J., Van Der Steen, N., Peeters, M., Rolfo, C., Deschoolmeester, V., and Pauwels, P. (2016). APR-246 (PRIMA-1(MET)) strongly synergizes with AZD2281 (olaparib) induced PARP inhibition to induce apoptosis in non-small cell lung cancer cell lines. Cancer Lett. *375*, 313–322. https://doi.org/10.1016/j.canlet.2016.03.017.
11. Bykov, V.J.N., Zache, N., Stridh, H., Westman, J., Bergman, J., Selivanova, G., and Wiman, K.G. (2005). PRIMA-1(MET) synergizes with cisplatin to induce tumor cell apoptosis. Oncogene *24*, 3484–3491. https://doi.org/10.1038/sj.onc.1208419.
12. Mohell, N., Alfredsson, J., Fransson, Å., Uustalu, M., Byström, S., Gullbo, J., Hallberg, A., Bykov, V.J.N., Björklund, U., and Wiman, K.G. (2015). APR-246 overcomes resistance to cisplatin and doxorubicin in ovarian cancer cells. Cell Death Dis. *6*, e1794. https://doi.org/10.1038/cddis.2015.143.
13. Liu, D.S., Duong, C.P., Haupt, S., Montgomery, K.G., House, C.M., Azar, W.J., Pearson, H.B., Fisher, O.M., Read, M., Guerra, G.R., et al. (2017). Inhibiting the system x(C)(-)/glutathione axis selectively targets cancers with mutant-p53 accumulation. Nat. Commun. *8*, 14844. https://doi.org/10.1038/ncomms14844.
14. Ceder, S., Eriksson, S.E., Cheteh, E.H., Dawar, S., Corrales Benitez, M., Bykov, V.J.N., Fujihara, K.M., Grandin, M., Li, X., Ramm, S., et al. (2021). A thiol-bound drug reservoir enhances APR-246-induced mutant p53 tumor cell death. EMBO Mol. Med. *13*, e10852. https://doi.org/10.15252/emmm.201910852.
15. McCubrey, J.A., Abrams, S.L., Fitzgerald, T.L., Cocco, L., Martelli, A.M., Montalto, G., Cervello, M., Scalisi, A., Candido, S., Libra, M., and Steelman, L.S. (2015). Roles of signaling pathways in drug resistance, cancer initiating cells and cancer progression and metastasis. Adv. Biol. Regul. *57*, 75–101. https://doi.org/10.1016/j.jbior.2014.09.016.
16. Tran, K.A., Kondrashova, O., Bradley, A., Williams, E.D., Pearson, J.V., and Waddell, N. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Med. *13*, 152. https://doi.org/10.1186/s13073-021-00968-x.
17. McInnes, L., Healy, J., and Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at arXiv.
18. Zhang, M., Zhang, Y.Y., Chen, Y., Wang, J., Wang, Q., and Lu, H. (2021). TGF-beta Signaling and Resistance to Cancer Therapy. Front. Cell Dev. Biol. *9*, 786728. https://doi.org/10.3389/fcell.2021.786728.
19. Elston, R., and Inman, G.J. (2012). Crosstalk between p53 and TGF-β Signalling. J. Signal Transduct. *2012*, 294097–294110. https://doi.org/10.1155/2012/294097.
20. Bugide, S., Parajuli, K.R., Chava, S., Pattanayak, R., Manna, D.L.D., Shrestha, D., Yang, E.S., Cai, G., Johnson, D.B., and Gupta, R. (2020). Loss of HAT1 expression confers BRAFV600E inhibitor resistance to melanoma cells by activating MAPK signaling via IGF1R. Oncogenesis *9*, 44. https://doi.org/10.1038/s41389-020-0228-x.
21. Krepler, C., Xiao, M., Samanta, M., Vultur, A., Chen, H.Y., Brafford, P., Reyes-Uribe, P.I., Halloran, M., Chen, T., He, X., et al. (2016). Targeting Notch enhances the efficacy of ERK inhibitors in BRAF-V600E melanoma. Oncotarget *7*, 71211–71222. https://doi.org/10.18632/oncotarget.12078.
22. Farnsworth, D.A., Inoue, Y., Johnson, F.D., de Rappard-Yuswack, G., Lu, D., Shi, R., Ma, L.I.J., Mattar, M.S., Somwar, R., Ladanyi, M., et al. (2022). MEK inhibitor resistance in lung adenocarcinoma is associated with addiction to sustained ERK suppression. npj Precis. Oncol. *6*, 88. https://doi.org/10.1038/s41698-022-00328-x.
23. Shchors, K., Persson, A.I., Rostker, F., Tihan, T., Lyubynska, N., Li, N., Swigart, L.B., Berger, M.S., Hanahan, D., Weiss, W.A., and Evan, G.I. (2013). Using a preclinical mouse model of high-grade astrocytoma to optimize p53 restoration therapy. Proc. Natl. Acad. Sci. USA *110*, E1480–E1489. https://doi.org/10.1073/pnas.1219142110.
24. Marzec, K.A., Lin, M.Z., Martin, J.L., and Baxter, R.C. (2015). Involvement of p53 in insulin-like growth factor binding protein-3 regulation in the breast cancer cell response to DNA damage. Oncotarget *6*, 26583–26598. https://doi.org/10.18632/oncotarget.5612.
25. Friedl, P., Vischer, P., and Freyberg, M.A. (2002). The role of thrombospondin-1 in apoptosis. Cell. Mol. Life Sci. *59*, 1347–1357. https://doi.org/10.1007/s00018-002-8512-9.
26. Daubon, T., Léon, C., Clarke, K., Andrique, L., Salabert, L., Darbo, E., Pineau, R., Guérit, S., Maitre, M., Dedieu, S., et al. (2019). Deciphering the complex role of thrombospondin-1 in glioblastoma development. Nat. Commun. *10*, 1146. https://doi.org/10.1038/s41467-019-08480-y.
27. Hu, C., Wen, J., Gong, L., Chen, X., Wang, J., Hu, F., Zhou, Q., Liang, J., Wei, L., Shen, Y., and Zhang, W. (2017). Thrombospondin-1 promotes cell migration, invasion and lung metastasis of osteosarcoma through FAK dependent pathway. Oncotarget *8*, 75881–75892. https://doi.org/10.18632/oncotarget.17427.
28. Park, S.G., Lee, T., Kang, H.Y., Park, K., Cho, K.H., and Jung, G. (2006). The influence of the signal dynamics of activated form of IKK on NF-kappaB and anti-apoptotic gene expressions: a systems biology approach. FEBS Lett. *580*, 822–830. https://doi.org/10.1016/j.febslet.2006.01.004.
29. Lee, H.S., Hwang, C.Y., Shin, S.Y., Kwon, K.S., and Cho, K.H. (2014). MLK3 is part of a feedback mechanism that regulates different cellular responses to reactive oxygen species. Sci. Signal. *7*, ra52. https://doi.org/10.1126/scisignal.2005260.
30. Seo, C.H., Kim, J.R., Kim, M.S., and Cho, K.H. (2009). Hub genes with positive feedbacks function as master switches in developmental gene regulatory networks. Bioinformatics *25*, 1898–1904. https://doi.org/10.1093/bioinformatics/btp316.
31. Atanasova, V.S., Russell, R.J., Webster, T.G., Cao, Q., Agarwal, P., Lim, Y.Z., Krishnan, S.,

Fuentes, I., Guttmann-Gruber, C., McGrath, J.A., et al. (2019). Thrombospondin-1 Is a Major Activator of TGF-beta Signaling in Recessive Dystrophic Epidermolysis Bullosa Fibroblasts. J. Invest. Dermatol. *139*, 1497–1505.e5. https://doi.org/10.1016/j.jid.2019.01.011.

32. Huang, S., Chen, B., Humeres, C., Alex, L., Hanna, A., and Frangogiannis, N.G. (2020). The role of Smad2 and Smad3 in regulating homeostatic functions of fibroblasts in vitro and in adult mice. Biochim. Biophys. Acta Mol. Cell Res. *1867*, 118703. https://doi.org/10.1016/j.bbamcr.2020.118703.
33. Brunen, D., Willems, S.M., Kellner, U., Midgley, R., Simon, I., and Bernards, R. (2013). TGF-beta: an emerging player in drug resistance. Cell Cycle *12*, 2960–2968. https://doi.org/10.4161/cc.26034.
34. Soucheray, M., Capelletti, M., Pulido, I., Kuang, Y., Paweletz, C.P., Becker, J.H., Kikuchi, E., Xu, C., Patel, T.B., Al-Shahrour, F., et al. (2015). Intratumoral Heterogeneity in EGFR-Mutant NSCLC Results in Divergent Resistance Mechanisms in Response to EGFR Tyrosine Kinase Inhibition. Cancer Res. *75*, 4372–4383. https://doi.org/10.1158/0008-5472.CAN-15-0377.
35. Menichini, P., Monti, P., Speciale, A., Cutrona, G., Matis, S., Fais, F., Taiana, E., Neri, A., Bomben, R., Gentile, M., et al. (2021). Antitumor Effects of PRIMA-1 and PRIMA-1(Met) (APR246) in Hematological Malignancies: Still a Mutant P53-Dependent Affair? Cells *10*, 98. https://doi.org/10.3390/cells10010098.
36. Raghavendra, Capital, Sanjay, and Hbku. (2019). Deep Learning for Anomaly Detection: A Survey. Preprint at arXiv. https://doi.org/10.48550/arXiv.1901.03407.
37. Sreenath, S.N., Cho, K.H., and Wellstead, P. (2008). Modelling the dynamics of signalling pathways. Essays Biochem. *45*, 1–28. https://doi.org/10.1042/BSE0450001.
38. Joo, J.I., Zhou, J.X., Huang, S., and Cho, K.H. (2018). Determining Relative Dynamic Stability of Cell States Using Boolean Network Model. Sci. Rep. *8*, 12077. https://doi.org/10.1038/s41598-018-30544-0.
39. Cho, S.H., Park, S.M., Lee, H.S., Lee, H.Y., and Cho, K.H. (2016). Attractor landscape analysis of colorectal tumorigenesis and its reversion. BMC Syst. Biol. *10*, 96. https://doi.org/10.1186/s12918-016-0341-9.
40. Ian, Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Networks. Preprint at arXiv.
41. Arjovsky, M., Chintala, S., and Bottou, L.e. (2017). Wasserstein GAN. Preprint at arXiv. https://doi.org/10.48550/arXiv.1701.07875.
42. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. (2017). Improved Training of Wasserstein GANs. Preprint at arXiv. https://doi.org/10.48550/arXiv.1704.00028.
43. Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., and Schmidt-Erfurth, U. (2019). f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. Med. Image Anal. *54*, 30–44. https://doi.org/10.1016/j.media.2019.01.010.
44. Peter, W., Jessica, B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., et al. (2018). Relational inductive biases, deep learning and graph networks. Preprint at arXiv. https://doi.org/10.48550/arXiv.1806.01261.
45. Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., et al. (2016). A Landscape of Pharmacogenomic Interactions in Cancer. Cell *166*, 740–754. https://doi.org/10.1016/j.cell.2016.06.017.
46. Seashore-Ludlow, B., Rees, M.G., Cheah, J.H., Cokol, M., Price, E.V., Coletti, M.E., Jones, V., Bodycombe, N.E., Soule, C.K., Gould, J., et al. (2015). Harnessing connectivity in a large-scale small-molecule sensitivity dataset. Cancer Discov. *5*, 1210–1223.
47. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.
48. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30. https://doi.org/10.1093/nar/28.1.27.
49. Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. Protein Sci. *28*, 1947–1951. https://doi.org/10.1002/pro.3715.
50. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., and Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. *51*, D587–D592. https://doi.org/10.1093/nar/gkac963.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-THBS1	Santa Cruz	Cat# sc-59887; RRID:AB_793045
Anti-TGFβ1	Santa Cruz	Cat# sc-130348; RRID:AB_1567351
Anti-SMAD2/3	Santa Cruz	Cat# sc-133098; RRID:AB_2193048
Anti-phospho-SMAD2/3	Cell Signaling Technology	Cat# 8828; RRID:AB_2631089
Chemicals, peptides, and recombinant proteins		
Crystal violet solution	Sigma-Aldrich	Cat# HT90132-1L
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	Cat# D2650-100ML
APR-246	MedChemExpress (MCE)	Cat# HY-19980
TGF-β1 protein	MCE	Cat# HY-P7118
RPMI 1640	WelGENE	Cat# LM011-02A
DMEM	WelGENE	Cat# LM001-05
Fetal bovines serum (FBS)	WelGENE	Cat# PK004-07
Lipofectamine	Invitrogen	Cat# 11668019
Puromycin	Sigma-Aldrich	Cat# P8833-100MG
SYBR Master Mix	Genet Bio	Cat# Q-9200
Critical commercial assays		
RNA extraction kit	iNtRON Biotechnology	Cat# 17211
DiaStar RT kit	Solgent	Cat# DR23-R10K
Deposited data		
Gene expression data	GDSC2	https://www.cancerrxgene.org
Drug response data	GDSC2	https://www.cancerrxgene.org
Drug response data	CTRPv2	https://portals.broadinstitute.org/ctrp.v2.1
Pathway data	MSigDB	https://www.gsea-msigdb.org
Python code and data used in this study	This paper	GitHub - https://github.com/Soominll/AnoDAN
Experimental models: Cell lines		
Human lung cancer cell line: NCI-H1793	Korean Cell Line Bank (KCLB)	Cat# 91793
Human lung cancer cell line: NCI-H2009	KCLB	Cat# 92009
Human lung cancer cell line: NCI-H1792	ATCC	Cat# CRL-5895
Human lung cancer cell line: HCC827	KCLB	Cat# 70827
Oligonucleotides		
Primer for GAPDH, see Table 2	This paper	N/A
Primer for THBS1, see Table 2	This paper	N/A
Primer for TGFβ1, see Table 2	This paper	N/A
Recombinant DNA		
shTHBS1 sequences: TGACATCAGTGAGACCGATTT; GCTGGAAAGATTTCACTGCAT	Sigma-Aldrich	N/A
Overexpression vector for THBS1	VectorBuilder	Cat# pDNA(VB220725-1808zjh)
Control vector for overexpression: pLV-Puro-CMV>ORF_Stuffer	VectorBuilder	Cat# pDNA(VB900122-0485asj)

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: pLP1	Invitrogen	N/A
Plasmid: pLP2	Invitrogen	N/A
Plasmid: pLP/VSVG	Invitrogen	N/A
Software and algorithms		
IncuCyte ZOOM	Essen Bioscience	N/A
Veriti 96-well Thermal Cycler	Applied Biosystems	Cat# A48141
QuantStudio 5 real-time PCR system	Applied Biosystems	Cat# A34322
WGAN-GP	Gulrajani et al.[40–42]	N/A
f-AnoGAN	Schlegl et al.[43]	N/A
GNN	Peter et al.[44]	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kwang-Hyun Cho (ckh@kaist.ac.kr).

Materials availability

The study did not generate new unique reagents.

Data and code availability

- Processed data used in this article are available at (GitHub - https://github.com/SoominII/AnoDAN). All original data can be downloaded from GDSC (https://www.cancerrxgene.org/), CTRPv2(https://portals.broadinstitute.org/ctrp.v2.1/), and MSigDB (https://www.gsea-msigdb.org) databases as listed in the key resources table.
- All original code has been deposited at (GitHub - https://github.com/SoominII/AnoDAN) and is publicly available as of the date of publication.
- 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 culture

Human lung cancer cell lines, NCI-H1793, NCI-H2009, and HCC827 were purchased from the Korean Cell Line Bank (KCLB), and NCI-H1792 was purchased from the American Type Culture Collection (ATCC). NCI-H1793 and HCC827 were cultured in RPMI 1640 media (WelGENE) supplemented with 10% Fetal bovines serum (FBS; WelGENE) and antibiotics, which is comprised of 100 units/ml of penicillin, 100 μg/ml of streptomycin, and 0.25 μg/ml of Fungizone (Life Technologies Corp.). NCI-H1792 and NCI-H2009 were cultured in DMEM media (WelGENE) with the same supplements. All cells were cultured in an incubator with 5% CO2 at a temperature of 37°C.

Reagents

Crystal violet solution and dimethyl sulfoxide (DMSO) were obtained from Sigma-Aldrich (Saint Louis, MO, USA). APR-246 (HY-19980) and TGF-β1 protein (HY-P7118) were purchased from MedChemExpress (MCE). 25 mg of APR-246 was diluted to the final concentration of 10 mM and treated in the cells for 4–5 days 10 μg/ml of TGF-β1 was treated in the cells for 24 h.

METHOD DETAILS

Data preparation

We developed AnoDAN by utilizing both gene expression and pathway data. We could access drug response data of a p53 reactivator, APR-246, from NCI60, CTRPv2, and GDSC2 databases. Of these, we mainly utilized gene expression data from the Genomics of Drug Sensitivity in Cancer (GDSC, https://www.cancerrxgene.org/),[45] as it provides the most comprehensive data. In GDSC2, 1,556 genes are encompassed in all the pathways we used, and for APR-246, there are 968 total cell lines (413 – resistant, 313 – sensitive, 242 – intermediate). Cell line-drug combinations were provided with the half-maximal inhibitory concentration (IC50) value for each cell line. The IC50 values were log-transformed and normalized using *Z* score normalization (Z_lnIC50), which sets the mean of all values to 0 and the standard deviation of all values to 1. Z_lnIC50 greater than 0.5 was classified as resistant, and less than −0.5 was classified as sensitive. The same approach of data

preprocessing was applied to the trametinib dataset, which was used to test the reliability of our model. The total number of genes is the same as APR-246 dataset, and the total number of cell lines is 767 (302 – resistant, 144 – sensitive, 321 – intermediate).

The second most comprehensive APR-246 data from the Cancer Therapeutics Response Portal v2 (CTRPv2, https://portals.broadinstitute.org/ctrp.v2.1/)[46] were also employed to assess the robustness of our findings. The total number of genes for CTRPv2 is also 1,556 and the total number of cell lines is 210 (53 – resistant, 50 – sensitive, 107 – intermediate). In CTRPv2, we *Z* score normalized the area under the fitted dose-response curve (AUC) values, where scores above 0.5 were identified as resistant and below −0.5 as sensitive. Additionally, we obtained pathway data from the Molecular Signatures Database (MSigDB, https://www.gsea-msigdb.org),[47] and specifically acquired canonical pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG).[48–50] In total, 34 pathways that are related to cancer and signal transduction were used. Each pathway data includes the genes involved in the pathway and the interactions among the genes. The interaction types were transformed into one-hot encoded forms and used as inputs together with the gene expression data.

Architecture and training procedure of AnoDAN

The AnoDAN framework, which is depicted in Figure 1, is comprised of a generator, a discriminator, and an encoder, and draws inspiration from the f-AnoGAN approach used in image analysis,[43] although with a different structure and application. The generator and a discriminator are trained using the WGAN-GP[40–42] method. All the hyperparameters employed during training process are listed in Table S1. To begin with, the goal of the generator is to generate synthetic gene expression data that the discriminator, whose role is to accurately classify whether a given sample is real or generated, cannot distinguish from real gene expression data. It is constructed with three dense layers, incorporating layer normalization, Leaky ReLu activation functions, and dropout as shown in Figure S7A. When training, the loss function is:

$$L_{gen} = -D(G(z)) \quad \text{(Equation 1)}$$

where z is the latent variable.

Next, we incorporated GNN into the discriminator to analyze gene connections and classify samples at the pathway-level, requiring the generator to produce biologically accurate samples. Graphs representing biological pathways are used, with nodes representing genes and edges representing gene interactions. Any pathway data can be used, and here we employed graphs from KEGG pathways. Here, the GNN takes the gene expression value and interaction type as inputs for each gene and edge, respectively. The GNN aggregates the gene and interaction data iteratively and updates the graphical representation at each time step through the multilayer perceptron model.[44] Through this process, each gene gets positional information in the pathway, which provides more biological bias to the model. Following the graph networks, the discriminator model undergoes four dense layers, incorporating Leaky ReLu and ReLu activation functions (Figure S7B). This enables us to investigate anomalous genes at the pathway-level. The loss function for discriminator is expressed as:

$$L_{disc} = D(\tilde{x}) - D(x) + \lambda\left(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1\right)^2 \quad \text{(Equation 2)}$$

where $\tilde{x} = G(z)$, x represents real data, λ is a gradient penalty coefficient, ∇ denotes gradients, $\hat{x} = \epsilon x + (1-\epsilon)\tilde{x}$, ϵ is a random number sampled from a uniform distribution [0,1), and $\|\cdot\|^2$ is the sum of squared gradients. The discriminator undergoes training n_critic times while the generator undergoes training 1 time, as outlined in Table S1. In summary, the GAN model is trained exclusively on gene expression data from drug-sensitive cell lines to capture distinctive characteristics of these cell lines.

After training the GAN model, the encoder is also trained on drug-sensitive cell line data (Figure 2A). The main role of the encoder is to map the real sensitive cell line data to the latent space. Following the encoder mapping, the pre-trained generator then generates data from the latent space, which is evaluated by the discriminator to calculate loss by adding data reconstruction residuals and residuals in feature space populated by the discriminator. The encoder is trained using the f-AnoGAN method.[43] The loss function for encoder is defined as:

$$L_{enc} = \frac{1}{n} \times \|x - G(E(x))\|^2 + \frac{\kappa}{n_d} \times \|D(x) - D(G(E(x)))\|^2 \quad \text{(Equation 3)}$$

where n represents the number of genes in a sample, n_d is the number of samples, and κ is a weighting factor. This encoder architecture consists of three dense layers with layer normalization, Leaky ReLu activation function, and dropout (Figure S7C). After some iterative processes, the encoder eventually acquires characteristics specific to sensitive cell lines, enabling the identification of anomalous genes and pathways in resistant cell line data. Pathways enriched with top-scored genes are analyzed and top-scored genes in the top pathways are identified to be primarily responsible for drug resistance.

Anomality scoring

The anomality scoring process utilizes the pre-trained encoder and generator. Any cell line gene expression can be used as an input, and the anomaly score is calculated using the difference between real (input) and generated data. As demonstrated in Figure 2B, the sample-level anomaly score is determined by calculating the sum of the squared difference between the input and generated data:

$$A_{sample} = \frac{1}{n} \times \|x - G(E(x))\|^2 \quad \text{(Equation 4)}$$

where n is the number of genes in a sample, and $\|\cdot\|^2$ is the sum of squared difference between real and generated data. The gene-level score is determined by calculating the absolute difference for each gene between the input and the generated data:

$$A_{gene} = |x - G(E(x))| \quad \text{(Equation 5)}$$

The pathway-level score is composed of two factors: the proportion of the genes with top 1% gene-level score of each cell line included in each pathway and the average gene-level anomaly score for the top 1% genes. The average gene-level anomaly score is utilized as a weight, and the proportion is multiplied by this weight:

$$A_{pathway} = \frac{n_g}{n_p} \times w_g \quad \text{(Equation 6)}$$

where n_g denotes the number of top 1% genes in a pathway, n_p is the total number of genes in a pathway, and w_g represents the average gene-level anomaly score across the cell lines.

Virus production for knockdown and overexpression experiments

HEK 293T cell was used to transfect shRNA and overexpression vectors targeting THBS1, following the manufacturer's protocol. Two shRNA sequences (shTHBS1; TGACATCAGTGAGACCGATTT; GCTGGAAAGATTTCACTGCAT; Sigma-Aldrich) and one overexpression vector (pDNA(VB220725-1808zjh); VectorBuilder) were used. Control vector (pDNA(VB900122-0485asj); VectorBuilder) and scrambled and packaging mix (pLP1, pLP2, and pLP/VSVG) were used to produce lentivirus using Lipofectamine (Invitrogen). In NCI-H1792, NCI-H1793, NCI-H2009, and HCC827, transduction was performed using virus product and 4 μg/ml of polybrene (Sigma-Aldrich). Infected cells were selected using 1 μg/ml of puromycin for at least one week.

Cell growth analysis and crystal violet assay

Cells were seeded at 3-5 x 10^3 cells/well in a 96 well plate. Cells were then treated with the drugs (APR-246) after 24 h. Cell growth was recorded every 3 h using IncuCyte ZOOM (Essen Bioscience) for 4–5 days. Cell viability was analyzed by measuring cell confluence using IncuCyte ZOOM 2016A software. After 4–5 days, cells were stained with 1% (w/v) crystal violet (Sigma-Aldrich) for 30 min at room temperature and washed with distilled water.

Total RNA extraction and quantitative real-time PCR (qRT-PCR) analysis

Following the manufacturer's guidelines, total RNA was extracted from cells employing a total RNA extraction kit from iNtRON Biotechnology. Subsequently, it was treated with RNase-free DNase I from Thermo Fisher Scientific Inc. to eliminate genomic DNA. The synthesis of complementary DNA (cDNA) was carried out using the DiaStar RT kit (Solgent). Reverse transcription PCR (RT-PCR) was then performed utilizing the Veriti 96-well Thermal Cycler (Applied Biosystems). Quantitative real-time PCR (qRT-PCR) was performed with the primers and cDNA using QuantStudio 5 real-time PCR system from Applied Biosystems and SYBR Master Mix from Genet Bio. Sequences of qRT-PCR primers are listed in Table 2.

Western blot analysis

Cells were initially rinsed with phosphate-buffered saline (PBS) and then subjected to lysis using the mixture of lysis buffer (20 mM HEPES (pH7.2), 150 mM NaCl, 0.5% Triton X-100, 10% glycerol, 0.1% SDS) in addition to 0.1% protease inhibitor and phosphatase inhibitor cocktail (Thermo-scientific). This lysis process was carried out while maintaining the samples on ice. For immunoblotting, anti-THBS1 (sc-59887), anti-TGFβ1 (sc-130348), and anti-SMAD2/3 (sc-133098) were obtained from Santa Cruz Biotechnology Inc., and anti-phospho-SMAD2/3 (#8828) was purchased from Cell Signaling Technology Inc.

QUANTIFICATION AND STATISTICAL ANALYSIS

The AnoDAN framework is explained in detail in Figure 1, as well as the Method details (Architecture and training procedure of AnoDAN). Anomaly scores were computed following the guidelines outlined in the "Anomality scoring" part of the Method details. Microsoft Excel was utilized to generate and analyze bar graphs and cell viability curves. In the figures, statistically significant distinctions were denoted with p values obtained from the Student's t test. Error bars in the figures represent the mean ± standard deviation (SD). The significance levels were indicated as follows: * for $p \leq 0.05$, ** for $p \leq 0.01$, *** for $p \leq 0.001$, **** for $p \leq 0.0001$.