Attention-Enhanced Deep Generative Modeling Enables Interpretable Prediction of Cancer Drug Sensitivity

Shuangxia Ren SHR81@PITT.EDU
Department of Biomedical Informatics
University of Pittsburgh
Pittsburgh, PA, United States

Aodong Qiu QIUAODON@PITT.EDU
Department of Biomedical Informatics
University of Pittsburgh
Pittsburgh, PA, United States
School of Medicine, Tsinghua Medicine
Tsinghua University
Beijing, China

Mengyao Lu MY.LU@PITT.EDU
Department of Biomedical Informatics
University of Pittsburgh
Pittsburgh, PA, United States

Xinghua Lu XINGHUA@PITT.EDU
Department of Biomedical Informatics, Department of Pharmaceutical Sciences
University of Pittsburgh
Pittsburgh, PA, United States

Abstract

Predicting how tumors respond to specific drugs is a fundamental obstacle in precision oncology, largely because of the vast molecular diversity observed across cancers. Deep learning approaches have demonstrated potential for modeling complex relationships between molecular profiles and drug responses, yet their adoption in clinical settings has been hindered by limited transparency and poor generalization across different datasets. In this work, we introduce Residual Attention Variational Autoencoder with Elastic Net (ResAttnVAE-EN), a deep generative model augmented with attention mechanisms that combines somatic genomic alterations (SGAs) and transcriptomic profiles to learn biologically interpretable cellular representations for predicting drug sensitivity. Leveraging large-scale pharmacogenomic datasets from the Genomics of Drug Sensitivity in Cancer (GDSC) database, we show that ResAttnVAE-EN consistently surpasses standard variational autoencoder and regression-based benchmarks across a broad panel of therapeutic compounds. The attention layers highlight genomically coherent pathway-level drivers, and the learned latent spaces encode cellular states that generalize across multiple cancer types for response prediction. Notably, models derived from cell line experiments effectively distinguish survival trajectories and treatment outcomes in independent The Cancer Genome Atlas (TCGA) lung cancer patient cohorts. These findings position attention-augmented deep generative approaches as a reliable and interpretable framework for clinically translatable drug sensitivity modeling. Code is available at https://github.com/renshuangxia/ResAttnVAE.

1. Introduction

Selecting effective therapies for individual cancer patients is a pressing clinical need, yet predicting how a specific tumor will respond to a given drug remains remarkably difficult. Tumors exhibit extensive molecular heterogeneity—even patients whose cancers harbor the same driver mutations can experience vastly different treatment outcomes (Iorio et al., 2016; Garnett et al., 2012). This variability limits the reliability of mutation-based biomarkers alone and highlights the need for models that can integrate diverse molecular signals to inform treatment decisions (Marquart et al., 2018). For clinicians, more accurate drug sensitivity predictions could directly improve therapeutic selection, reduce exposure to ineffective regimens, and support more personalized care.

From a machine learning perspective, this prediction task is challenging because it requires capturing complex, nonlinear relationships among high-dimensional genomic and transcriptomic features while maintaining generalizability across biologically distinct cancer types. Traditional approaches based on linear or kernel-based models using gene expression or mutation data offer interpretability but fall short in modeling coordinated pathway activity and higher-order feature interactions (Geeleher et al., 2014; Ding et al., 2018). Deep learning methods, particularly deep generative models such as variational autoencoders (VAEs) (Kingma and Welling, 2014), can learn compact latent representations that summarize rich molecular profiles and encode cellular states spanning tissue boundaries (Baptista et al., 2021; Way and Greene, 2018). However, these models frequently sacrifice interpretability—a critical requirement for clinical trust and adoption—making it difficult for practitioners to understand why a particular prediction is made.

Several efforts have sought to address these limitations. Prior work has explored supervised deep learning architectures for drug response modeling (Menden et al., 2013; Chiu et al., 2019; Li et al., 2021), as well as generative approaches that learn biologically meaningful embeddings from omics data (Dincer et al., 2020; Xue et al., 2020). Interpretable architectures such as DrugCell (Kuenzi et al., 2020) embed biological ontologies into neural networks to provide mechanistic insight. Attention mechanisms (Vaswani et al., 2017), widely successful in natural language processing and computer vision, offer a principled means of weighting input features by relevance and have begun to be applied in biomedical contexts (Tao et al., 2020). Nevertheless, the integration of attention into deep generative frameworks for pharmacogenomic prediction remains largely unexplored, leaving a gap between predictive performance and mechanistic transparency.

In this work, we introduce Residual Attention Variational Autoencoder with Elastic Net (ResAttnVAE-EN), a two-stage framework that pairs a residual attention-based variational autoencoder (ResAttnVAE) with Elastic Net (EN) drug response classifiers (Fig. 1). ResAttnVAE incorporates somatic genomic alterations (SGAs) and cancer type information into multiple hidden layers through self-attention and residual connections, enabling the model to learn pathway-informed cellular representations. These representations are then used by EN classifiers for drug sensitivity prediction, combining robust performance with feature-level interpretability. We benchmark our approach against multiple generative and end-to-end baselines, examine the biological coherence of attention-derived insights, and validate clinical relevance using independent patient cohorts.

This work makes four contributions: (i) attention within a VAE's residual pathway for tumor-specific SGA weighting, unexplored in pharmacogenomics; (ii) a two-stage design outperforming an end-to-end variant with the same backbone (AUROC 0.719 vs. 0.665); (iii) interpretability without sacrificing performance; and (iv) integrating the full genomic/transcriptomic profile, unlike single-gene testing, to catch responsive patients biomarker testing would miss.

Generalizable Insights about Machine Learning in the Context of Healthcare

This work offers several insights for the machine learning in healthcare community. First, incorporating attention mechanisms into deep generative models yields both improved predictive accuracy and interpretable, pathway-coherent feature attributions—suggesting that interpretability and performance need not be at odds. Second, latent representations learned from cell line pharmacogenomic data transfer meaningfully to patient settings, stratifying survival and treatment response in independent clinical cohorts. Finally, the two-stage design—separating representation learning from prediction—provides a modular paradigm adaptable to other healthcare domains where integrating heterogeneous data and maintaining transparency are equally important.

2. Related Work

Deep learning and representation learning for drug response. Deep learning methods have been increasingly applied to drug response prediction (Baptista et al., 2021). Menden et al. (2013) were among the first to use neural networks for this task, and more recent architectures include deep neural networks on integrated genomic profiles (Chiu et al., 2019) and DeepDSC (Li et al., 2021). However, most models treat prediction as a single end-to-end task, risking overfitting when labeled data are limited (Costello et al., 2014). An alternative strategy first learns meaningful representations and then uses them for downstream prediction. VAEs (Kingma and Welling, 2014) have been adopted to learn compact latent spaces from molecular profiles (Way and Greene, 2018). Dincer et al. (2020) used VAEs for drug response prediction from gene expression embeddings, and Xue et al. (2020) trained deep generative models on transcriptomic perturbation data.

Interpretable models and attention mechanisms. A persistent challenge in clinical deep learning is insufficient interpretability. Kuenzi et al. (2020) addressed this with DrugCell, embedding the Gene Ontology hierarchy into a neural network. Pathway-guided approaches such as PathDNN (Deng et al., 2020) incorporate known biological pathways into network architectures. Attention mechanisms (Vaswani et al., 2017) offer another route to interpretability, and Tao et al. (2020) introduced contextual attention for drug sensitivity prediction. Despite these advances, the integration of attention into deep generative models for pharmacogenomic prediction remains largely unexplored.

Clinical translation from cell line models. A critical question for any pharmacogenomic model is whether predictions trained on cell line data generalize to patient settings. Several studies have demonstrated that cell-line-derived models can stratify patient outcomes (Geeleher et al., 2014; Ding et al., 2018), though challenges remain due to distributional differences between preclinical and clinical data. Prior work showed that both

DeepGRMF (Ren et al., 2022) and ResGitDR (Ren et al., 2024) predictions correlate with survival outcomes in TCGA lung cancer patients, supporting the feasibility of cell-line-to-patient transfer for drug sensitivity models.

Positioning of the present work. ResAttnVAE-EN integrates attention mechanisms into a VAE framework for pharmacogenomic representation learning. Unlike DrugCell, it learns pathway-coherent representations without explicit ontology structure. Unlike end-to-end approaches, it separates representation learning from prediction, improving stability. The attention mechanism provides genomic-level interpretability, while Elastic Net classifiers support feature-level transparency.

3. Methods

3.1. Model Architecture and Training

Residual Attention Variational Autoencoder with Elastic Net (ResAttnVAE-EN) is a two-stage framework consisting of a representation learning module and a drug response prediction module (Fig. 1).

Representation Learning Module (Fig. 1A). The representation learning component employs a VAE that reconstructs gene expression profiles while integrating SGA information through self-attention and residual connections. Each SGA gene is associated with two learnable embedding vectors, $\mathbf{e}_m^1$ and $\mathbf{e}_m^2$, capturing its effects at different network depths: $\mathbf{e}_m^1$ operates at the encoder and decoder hidden layers, while $\mathbf{e}_m^2$ acts at the latent bottleneck. Embeddings are initialized using Gene2Vec pretrained representations (Ren et al., 2024; Du et al., 2019) (see Appendix B).

Attention Mechanism for SGA Weighting. A multi-head self-attention mechanism computes tumor-specific importance weights over SGA embeddings (Tao et al., 2020), enabling the model to dynamically prioritize different genomic alterations depending on the mutational context of each tumor (Fig. 1A).

Residual Integration and Latent Representation Learning. Attention-weighted signal embeddings are incorporated into the VAE through residual connections (Fig. 1A). Two signal embeddings are computed from the two attention blocks:

$$e_1 = \sum_m \alpha_m^1 \cdot e_m^1 \tag{1}$$

$$e_2 = \sum_m \alpha_m^2 \cdot e_m^2 \tag{2}$$

where α_m^1 and α_m^2 are the attention weights for SGA gene m from the first and second attention blocks, respectively. The encoder hidden representation is obtained by combining the input gene expression X with signal embedding e_1 through a residual connection:

$$H_1 = \mathrm{ReLU}(W_1 \cdot X + e_1) \tag{3}$$

At the latent bottleneck, signal embedding e_2 is integrated to obtain the variational parameters:

$$\mu = W_\mu(H_1 + e_2) \tag{4}$$

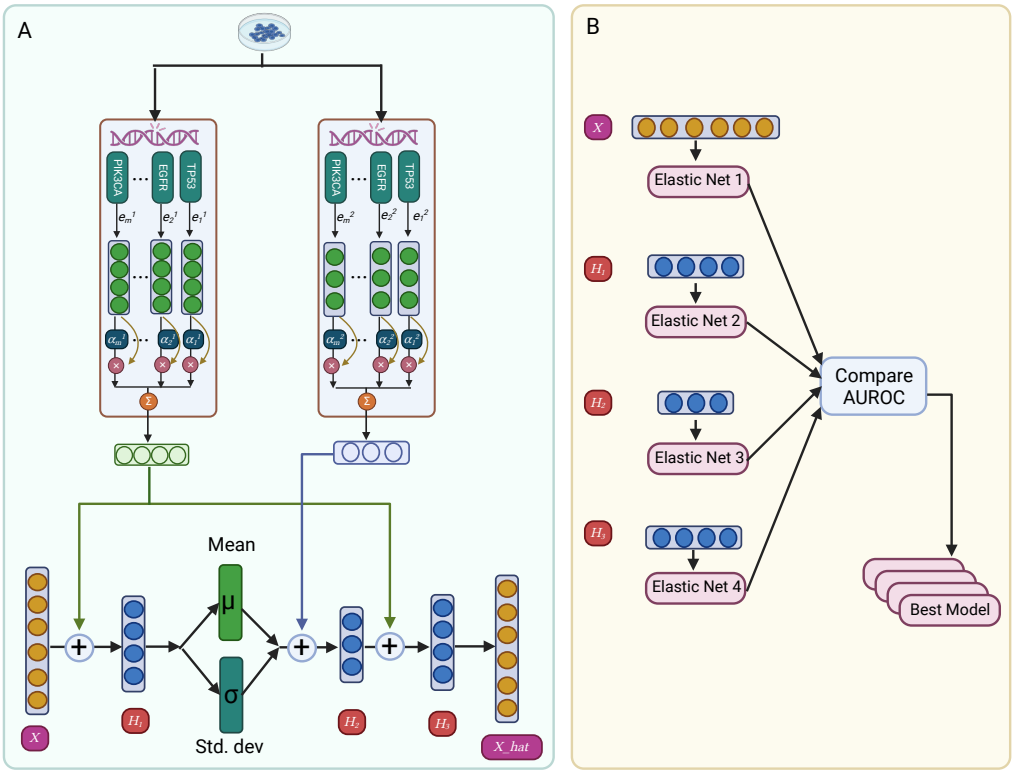

Figure 1: The architecture of ResAttnVAE-EN. (A) The variational autoencoder reconstructs gene expression while integrating SGA information through self-attention and residual connections. (B) Original gene expression and each hidden representation are used to predict drug response with Elastic Net models; the best-performing model is selected per drug.

$$\log \sigma^2 = W_\sigma(H_1 + e_2) \tag{5}$$

$$z = \mu + \sigma \cdot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \tag{6}$$

The decoder mirrors the encoder structure, with signal embedding e_1 added through a residual connection to produce the decoded hidden representation:

$$H_3 = \mathrm{ReLU}(W_3 \cdot z + e_1) \tag{7}$$

$$\hat{X} = W_{out} \cdot H_3 \tag{8}$$

The sampled latent variable z serves as the second hidden representation H_2. The training objective combines a mean squared error reconstruction loss with a Kullback–Leibler divergence term regularizing the latent distribution toward a standard normal prior:

$$\mathcal{L} = \mathrm{MSE}(X, \hat{X}) + D_{KL}(q(z|X) \| p(z)) \tag{9}$$

Drug Response Prediction Module (Fig. 1B). Latent representations extracted from each hidden layer of ResAttnVAE, along with original gene expression profiles, serve as candidate feature sets for drug response prediction. Elastic Net (EN) logistic regression was selected as the downstream classifier (Zou and Hastie, 2005) due to its capacity for sparse feature selection while accommodating correlated predictors. For each compound, EN models are trained separately on four feature sets: raw gene expression, first hidden layer representation, second hidden layer representation, and third hidden layer representation. Hyperparameters C and L1_ratio are optimized via grid search with 10-fold cross-validation. For each drug, the representation yielding the highest average AUROC is selected as the final model (Fig. 1B). Identical cross-validation partitions are enforced across both representation learning and supervised prediction stages to prevent information leakage.

Baseline Architectures (Fig. 2). We compare ResAttnVAE-EN against several baselines (Fig. 2). The standalone Elastic Net (EN) is the simplest baseline, training directly on raw SGAs and gene expression data without any representation learning.

The remaining baselines (VAE-EN, ResVAE-EN, and ResAttnVAE-EN) all follow the same two-stage design: a representation learning module is first trained to learn hidden representations of cellular states, which are then used as input features to a separate Elastic Net classifier for each drug. These models differ only in how the representation learning module is constructed. VAE-EN (Fig. 2A) uses a standard VAE that concatenates gene expression and SGAs as a single input vector, without further genomic integration at deeper layers. ResVAE-EN (Fig. 2B) uses gene expression as both input and reconstruction target, and injects transformed SGA vectors into each hidden layer via feedforward residual connections, but treats all SGAs equally without tumor-specific weighting. ResAttnVAE-EN extends this by replacing the feedforward injection with a multi-head self-attention mechanism, enabling the model to dynamically assign tumor-specific importance weights to SGAs before integrating them into the hidden layers.

Finally, ResAttnVAEDR (Fig. 2C) shares the same attention-enhanced representation learning as ResAttnVAE-EN, but collapses the two-stage design into a single end-to-end network that jointly optimizes representation learning and drug response prediction in one objective.

Model Complexity. ResAttnVAE-EN has more parameters than EN and VAE-EN (SGA embeddings, Gene2Vec, attention), added for interpretability rather than raw performance. That complexity maintains rather than degrades performance suggests attention captures real biological structure; we hypothesize this advantage grows with more data (a scaling-law effect), left for future work.

Clinical Prediction and Survival Analysis. For clinical validation, ResAttnVAE-EN was evaluated in two complementary settings, both following previously established frameworks (Ren et al., 2022, 2024).

Survival analysis. Clinical data including treatments and overall survival were collected for 62 LUAD and LUSC patients from TCGA who received combinations of cisplatin, pemetrexed, paclitaxel, and/or vinorelbine as adjuvant therapies. EN models trained on GDSC data were applied to derive sensitivity probabilities, and patients were stratified into predicted responders and non-responders using a percentile-based threshold. Overall

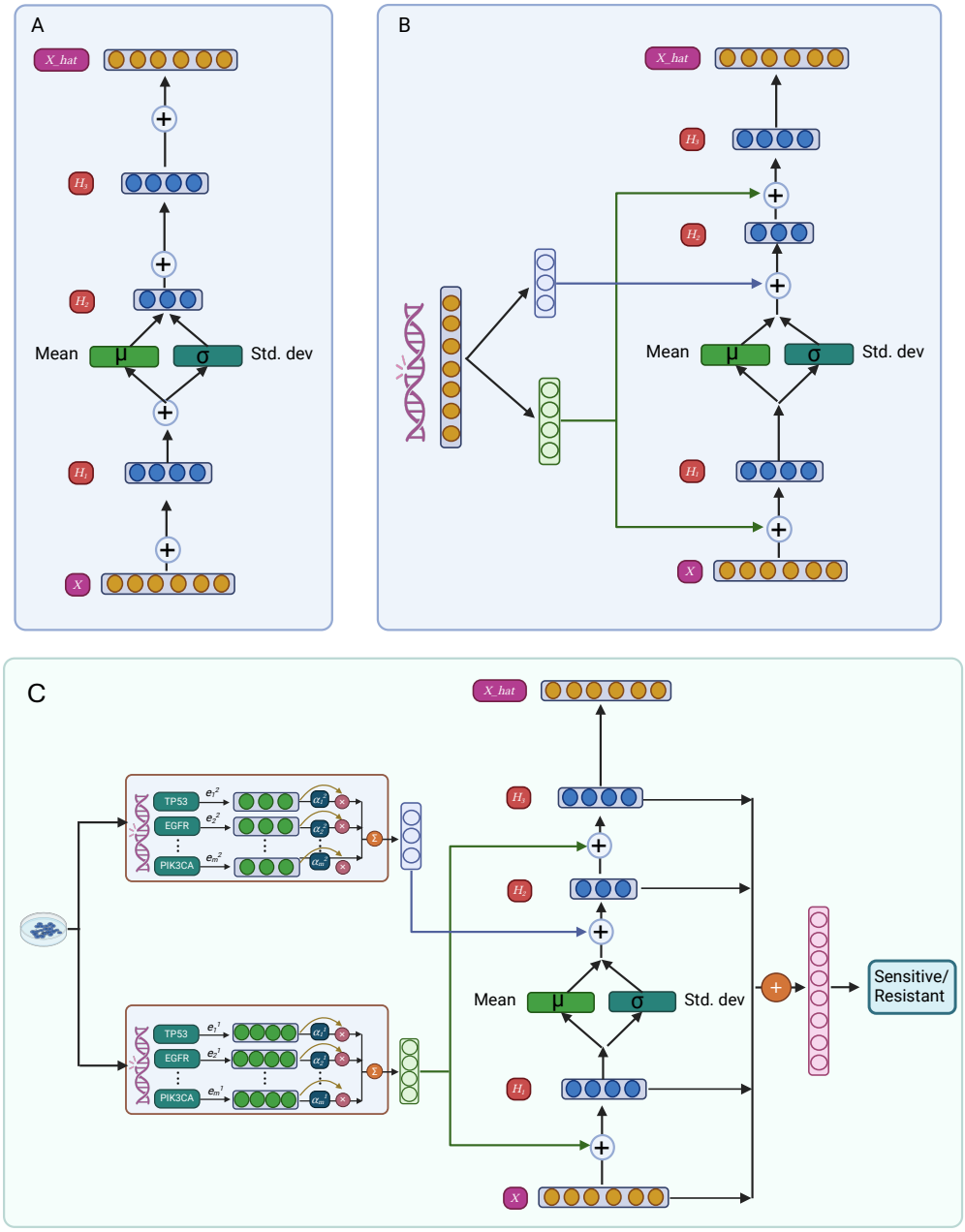

Figure 2: Baseline architectures. (A) Standard VAE concatenates all inputs. (B) ResVAE adds SGA vectors via residual connections. (C) ResAttnVAEDR uses attention-enhanced representations to predict drug response end-to-end.

survival differences were assessed using Kaplan–Meier curves with log-rank testing (see Appendix D for details).

Drug response prediction in TCGA. Predictive accuracy was further evaluated using curated TCGA drug response data from Ding et al. (2016), comprising 1,133 records across 762 patients after filtering. For patients receiving multiple drugs, overall sensitivity was computed using a noisy-or aggregation. A drug coverage ratio was introduced to assess how prediction accuracy scales with the completeness of pharmacological information (see Appendix D).

Embedding Analysis Using Connectivity Maps. To assess whether learned representations capture biologically meaningful structure, connectivity maps were constructed for both SGA genes and drugs based on cosine similarity of their learned embeddings (Ren et al., 2024) (see Appendix E for construction details).

4. Cohort

4.1. Cohort Selection

Pharmacogenomic and molecular data were prepared following previously described procedures (Ren et al., 2024). Cancer cell line data were obtained from the GDSC project (Yang et al., 2013), with cell lines included if they possessed complete gene expression profiles, SGAs, and dose–response measurements. For clinical validation, two independent TCGA (Tomczak et al., 2015) patient cohorts were used: (1) for survival analysis, LUAD and LUSC patients with RNA-seq profiles, documented chemotherapy regimens, and recorded overall survival, restricted to those receiving adjuvant chemotherapy; and (2) for drug response evaluation, curated pan-cancer patient–drug response annotations from Ding et al. (2016), dichotomized into responders and non-responders (Geeleher et al., 2014).

4.2. Data Extraction

Somatic Genomic Alterations. SGA preprocessing followed previously described procedures (Ren et al., 2024). Each SGA was represented as a binary variable indicating the presence of somatic mutations or copy number alterations. The analysis was restricted to a curated panel of 1,084 genes compiled from three complementary sources: 527 driver genes from the Cell Model Passports, 634 expression-regulatory genes identified by Cai et al. (2020), and 324 mutation-associated genes used by Foundation Medicine (see Appendix C for data sources and preprocessing details).

Gene Expression. Gene expression data for GDSC cell lines consisted of RMA-normalized microarray measurements (Garcia-Alonso et al., 2018), while TCGA expression profiles were derived from RNA sequencing. Genes were selected using variance-driven criteria described by Ding et al. (2018). Cross-platform harmonization was achieved through nonparanormal (NPN) transformation followed by min–max normalization, enabling transfer of cell-line-trained models to patient cohorts without shared-sample batch correction (see Appendix C).

Drug Sensitivity. Drug sensitivity measurements were obtained from the GDSC website and summarized using activity area (AA) across 367 compounds. Responses were discretized into sensitive and resistant categories using the waterfall method (Ding et al., 2018) (see Appendix C for binarization details). This procedure yields a median sensitive fraction of approximately 15.5% across the 367 drugs, reflecting realistic pharmacological response

Table 1: Comparative performance of various deep generative models for predicting drug sensitivity. ResAttnVAE-EN achieves the highest average AUROC and the largest number of well-predicted drugs across all thresholds.

Method	Average AUROC	# AUROC > 0.7	# AUROC > 0.8	# AUROC > 0.9
EN	0.7089	193	58	2
VAE-EN	0.7180	219	61	**3**
ResVAE-EN	0.7182	213	62	**3**
ResAttnVAE-EN	**0.7190**	**221**	**65**	**3**
ResAttnVAEDR	0.6647	119	21	0

rates rather than a balanced design choice. Because AUROC is threshold-independent and robust to class imbalance, we adopt it as our primary evaluation metric throughout.

5. Results on Real Data

5.1. Evaluation Approach

All models were assessed using 10-fold cross-validation on the GDSC dataset spanning 367 compounds. Predictive performance was measured by area under the receiver operating characteristic curve (AUROC), and models were compared across multiple clinically meaningful thresholds. Clinical generalizability was evaluated by applying cell-line-trained models to independent TCGA lung cancer patient cohorts to assess survival stratification and treatment response prediction.

5.2. ResAttnVAE-EN Outperforms Baseline Approaches in Drug Sensitivity Prediction

As reported in Table 1, ResAttnVAE-EN achieved the highest average AUROC (0.7190) among all evaluated models and identified the largest number of compounds with AUROC values surpassing the 0.7 and 0.8 thresholds (221 and 65, respectively). Although VAE-EN and ResVAE-EN attained comparable average AUROC values, the key advantage of ResAttnVAE-EN lies not in average AUROC alone but in consistently identifying a broader set of well-predicted drugs, indicating improved robustness across the compound library rather than gains concentrated among a few drugs.

End-to-end deep learning architectures (ResAttnVAEDR) performed substantially worse, with markedly lower average AUROC (0.6647 vs. 0.7190) and fewer compounds exceeding any performance threshold. This sharp performance drop, despite ResAttnVAEDR sharing the same attention-enhanced representation learning as ResAttnVAE-EN, further confirms that the two-stage design—rather than model capacity alone—prevents overfitting. These findings suggest that jointly optimizing representation learning and response prediction in a single supervised objective may be prone to overfitting in pharmacogenomic settings where labeled training examples are scarce. Decoupling unsupervised representation learning from downstream prediction, as implemented in ResAttnVAE-EN, produces more stable and transferable models.

Table 2: The number of drugs achieving the highest AUROC using original gene expression or various hidden representations. Different drugs preferentially rely on different levels of abstraction.

Method	Original gene expression	First hidden	Second hidden	Third hidden
VAE-EN	200	155	3	9
ResVAE-EN	219	128	4	16
ResAttnVAE-EN	195	152	5	15

5.3. Latent Representations Complement Raw Gene Expression for Drug Response Prediction

We investigated which feature sets contributed most strongly to prediction by examining, for each drug, whether raw gene expression or a particular hidden layer representation yielded the highest AUROC (Table 2). While original expression profiles remained the top-performing feature set for a sizable proportion of compounds, latent representations—especially those from early hidden layers—frequently surpassed raw features in ResAttnVAE-EN.

This pattern carries two implications: representations from ResAttnVAE encode biological signal not readily captured by raw expression, and different compounds favor different levels of abstraction—some driven by a few informative genes captured directly by EN, others by pathway-level activity better summarized by SGA-guided latent representations. Since the drug response module selects the best feature set per compound (Section 3.1), the framework adapts to both.

5.4. SGA Embeddings Reflect Pathway-Level Organization

In ResAttnVAE, each SGA gene is associated with two learnable embedding vectors capturing its impact on the cellular signaling system at different levels of the network hierarchy. The model learns these embeddings solely through gene expression reconstruction, without receiving any explicit pathway annotations. Under this design, SGAs that perturb distinct members of a common signaling pathway are expected to exert similar functional impacts and thus develop similar embeddings. To test this, we computed pairwise cosine similarities across all SGA embeddings and identified the top 10 nearest neighbors for each gene.

We used the ten major cancer pathways reported by Sanchez-Vega et al. (2018) as ground truth for evaluating pathway coherence, constructing a nearest-neighbor connectivity map among SGA genes shared between the training dataset and the reported pathway gene sets. An edge was drawn between two genes if one or both ranked among the other's top neighbors, with edges colored by pathway membership (Fig. 3A). The embeddings of PI3K pathway members — *PIK3CA*, *PIK3R1*, and *PTEN* — consistently ranked among each other's closest neighbors, and comparable clustering was observed across other major pathways including RTK–RAS and HIPPO. These results indicate that ResAttnVAE acquires structured genomic representations aligned with established pathway organization, emerging purely from the supervised signal of gene expression prediction.

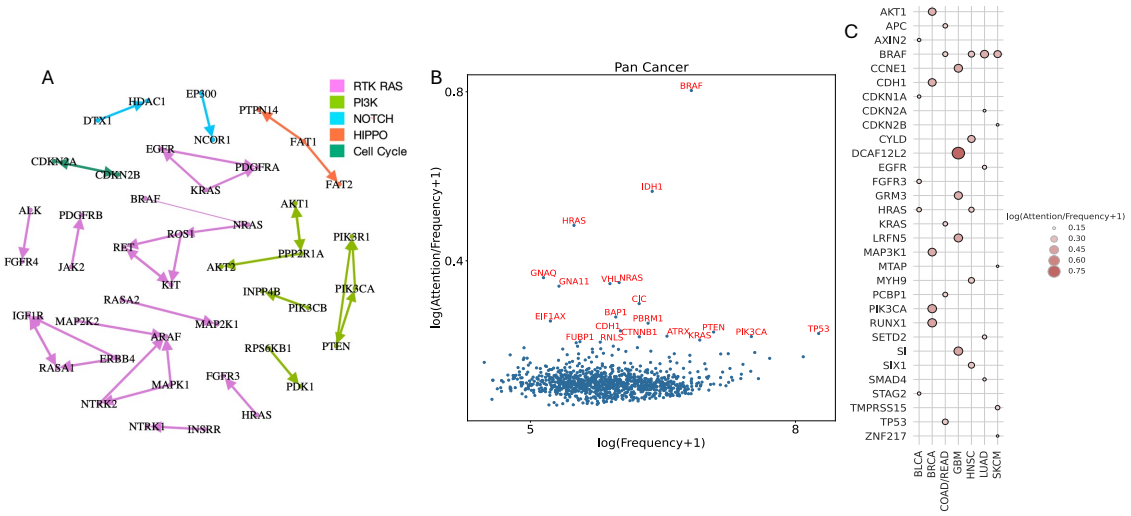

Figure 3: (A) Connectivity map illustrating cosine similarity among SGA gene embeddings; genes affecting shared pathways cluster together. (B) Pan-cancer attention weights assigned to SGA genes, with well-known oncogenic drivers highlighted in red. (C) Cancer-type-specific attention weight distributions across individual cancer types.

5.5. Attention Weights Highlight Biologically Coherent Genomic Drivers

ResAttnVAE employs a multi-head self-attention mechanism to assign tumor-specific importance weights to each SGA, reflecting its relative contribution to the reconstruction of gene expression. We aggregated these weights across all cell lines to examine which SGAs were broadly influential in pan-cancer settings (Fig. 3B) and which were specifically important in individual cancer types (Fig. 3C).

As shown in Fig. 3B, ResAttnVAE assigned high attention weights to well-characterized oncogenic drivers, such as *BRAF*, *KRAS*, *NRAS*, *HRAS*, and *PIK3CA*. Interestingly, several genes encoding signaling proteins — including *GNAQ* and *GNA11* — also received high attention weights despite their relatively low mutation frequency. These genes have not traditionally been considered major cancer drivers, yet recent research has highlighted their significant roles in tumorigenesis (Raamsdonk et al., 2010), suggesting that the model captures functional relevance rather than simply reflecting mutation prevalence. Individual attention weights vary somewhat across seeds, but canonical drivers (*BRAF*, *KRAS*, *PIK3CA*) consistently score high, so we emphasize the pattern's biological coherence over exact reproducibility.

Cancer-type-specific attention analyses further demonstrated context-dependent SGA prioritization (Fig. 3C). For example, *PIK3CA* and *MAP3K1* received prominent attention weights in breast invasive carcinoma (BRCA), consistent with their established roles in this disease, while *EGFR* and *BRAF* were most highly weighted in lung adenocarcinoma (LUAD), reflecting known tissue-specific oncogenic dependencies. In bladder urothelial

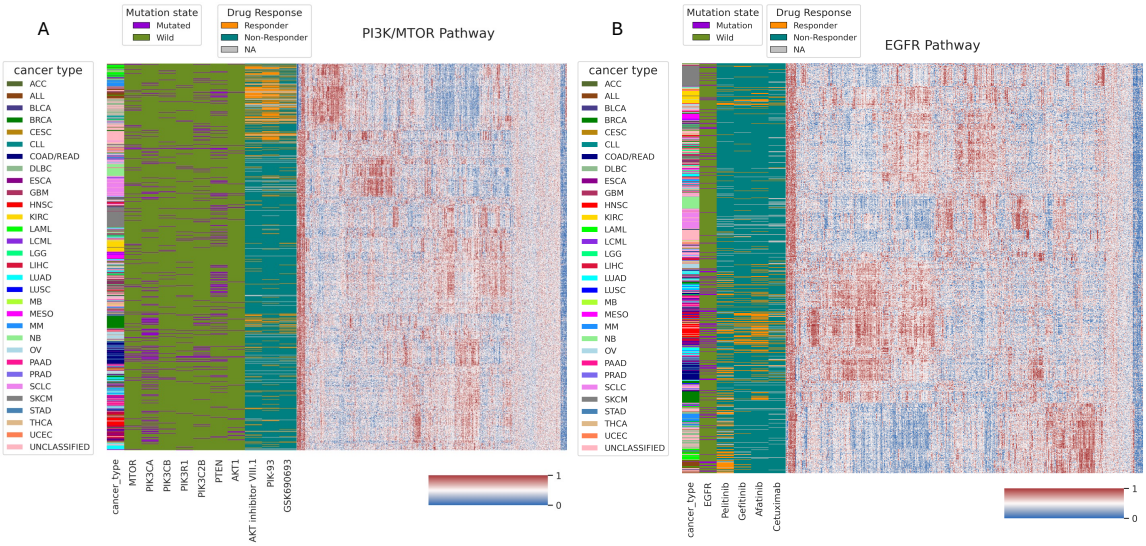

Figure 4: Cell-state-oriented prediction of drug sensitivity using ResAttnVAE-EN. (A) Cancer cell lines clustered based on top predictive features for PI3K/mTOR pathway drugs, with binary drug responses and mutation statuses of pathway genes displayed. (B) Clustering for EGFR-targeted drugs with EGFR mutation status annotated. Responder-enriched clusters span multiple cancer types in both panels.

carcinoma (BLCA) and head and neck squamous cell carcinoma (HNSC), distinct SGA prioritization patterns were also observed, as were tissue-specific weightings in colon and rectal adenocarcinoma (COAD/READ), lung squamous cell carcinoma (LUSC), and skin cutaneous melanoma (SKCM).

5.6. Cell-State Representations Capture Cross-Cancer-Type Drug Sensitivity Patterns

Contemporary genome-informed precision oncology assigns treatment based on the genomic status of individual signaling proteins. To evaluate whether inferred cellular states offer advantages over conventional genomic biomarkers, we examined drugs targeting the PI3K/mTOR pathway by testing whether cell lines predicted to be sensitive by ResAttnVAE-EN exhibit distinct response patterns compared to those stratified by individual pathway SGAs alone. We extracted the top predictive features for three PI3K/mTOR-targeting drugs — AKT inhibitor VIII, PIK-93, and GSK690693 — based on absolute Elastic Net weights, and grouped GDSC cell lines using clustering analysis (Fig. 4A). The mutation status of PI3K/mTOR pathway genes is overlaid to assess whether individual SGAs serve as informative biomarkers. The figure shows that inferred cell states form clusters spanning diverse cancer types, with certain clusters enriched with responders to all three drugs,

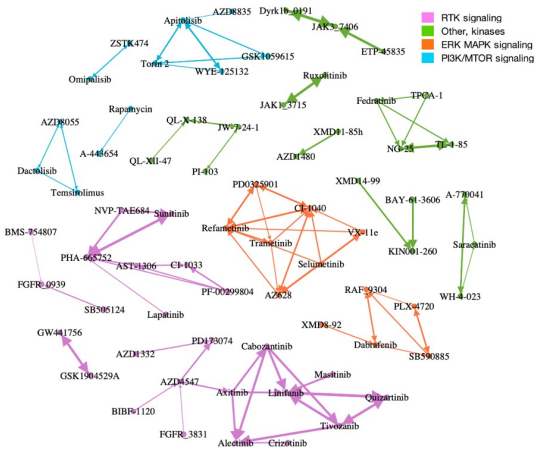

Figure 5: Connectivity map of drug embeddings derived from Elastic Net coefficients. Drugs targeting shared molecular pathways cluster together based on cosine similarity, with edge thickness corresponding to similarity strength.

suggesting that shared cellular states underlie sensitivity across this drug class rather than individual genomic biomarkers alone.

Analogous patterns emerged for EGFR-targeted agents (Fig. 4B), where multiple responder-enriched clusters were identified across diverse tissue origins, with enrichment linked to EGFR pathway mutation status. Importantly, tissue of origin alone did not account for the observed clustering, supporting the view that inferred cellular states offer a more informative basis for drug sensitivity prediction than cancer type classification alone. These findings indicate that ResAttnVAE-EN predicts drug response in a cell-state-oriented manner rather than relying on the genomic status of individual biomarker genes.

Notably, mutation status of individual pathway genes (e.g., *PIK3CA*, *PTEN*, *EGFR*) is not uniformly positive within these clusters—many responders are wild-type—quantifying the limitation of single-gene testing that cell-state predictions can overcome.

5.7. Drug Embeddings Capture Shared Mechanisms of Action

We further investigated ResAttnVAE-EN from a drug-centered perspective. For each drug, we extracted the Elastic Net coefficient vectors as "drug embeddings" and performed pairwise cosine similarity analysis, visualizing the five most similar drugs per compound in a connectivity map (Fig. 5; see Appendix E for details). Drugs targeting common signaling pathways — including RTK, ERK MAPK, and PI3K/mTOR signaling — consistently shared similar embeddings, supporting the interpretation that ResAttnVAE-EN identifies cellular state features indicative of sensitivity to drugs sharing mechanisms of action.

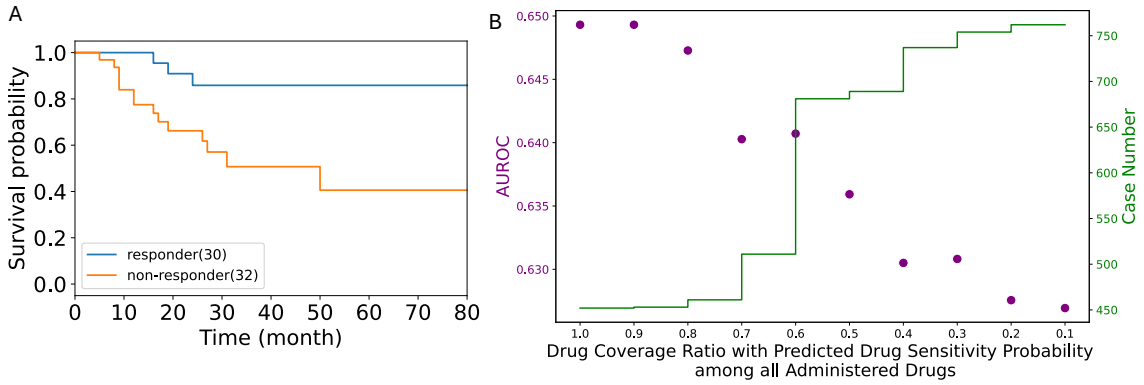

Figure 6: Clinical validation in TCGA lung cancer patients. (A) Kaplan–Meier survival curves for predicted responder and non-responder groups based on ResAttnVAE-EN predictions for patients receiving cisplatin, pemetrexed, paclitaxel, and/or vinorelbine as adjuvant therapy. (B) AUROC for predicting drug response at varying drug coverage ratio thresholds.

5.8. Cell-Line-Trained Models Stratify Survival and Treatment Response in TCGA Lung Cancer Patients

The above results support the effectiveness of ResAttnVAE-EN in predicting drug sensitivity in cell lines. We also assessed real-world value via two evaluations: a smaller exploratory survival analysis, and a larger drug-response cohort.

Survival analysis (exploratory). Clinical data including treatments and overall survival were collected for 62 LUAD/LUSC TCGA patients treated with cisplatin, pemetrexed, paclitaxel, and/or vinorelbine. This small, retrospective cohort is treated as exploratory—showing correlation with real-world outcomes rather than definitive validation. GDSC-trained EN models classified patients as predicted responders (n=30) or non-responders (n=32) via a percentile threshold approximating the average FDA-approved-drug response rate (Appendix D). Kaplan–Meier analysis showed responders had significantly longer survival (Fig. 6A), indicating meaningful signal despite cell-line-to-patient differences.

TCGA drug response (broader grounding). Using curated TCGA annotations (Ding et al., 2016) (1,133 records, 762 patients), 27 shared drugs, and noisy-or aggregation for combination regimens, AUROC across drug coverage thresholds (Fig. 6B; Appendix D) reached ∼0.65 at full coverage, decreasing with lower coverage. Given only 27/367 drug overlap and no shared calibration samples, an above-chance AUROC is consistent with prior transfer studies (Geeleher et al., 2014; Ding et al., 2018) and indicates real-world signal, not clinical-grade accuracy. Together, both evaluations provide converging, preliminary evidence of meaningful cell-to-patient transfer.

6. Discussion

This study presents ResAttnVAE-EN, an interpretable deep generative framework for drug sensitivity prediction that models pathway-informed cellular states through attention-enhanced residual learning. By jointly leveraging somatic genomic alterations, gene expression, and cancer type information, the approach addresses two persistent challenges in computational pharmacogenomics: limited model transparency and poor cross-dataset generalization.

The modest average AUROC gain over the strongest baselines is consistent with the design goal: prioritizing interpretability over raw performance, with maintained accuracy itself a central result. The clearer advantage is breadth—the most well-predicted drugs (221 above AUROC 0.7, 65 above 0.8)—indicating consistency across the compound library rather than gains concentrated in a few drugs.

A principal finding is that drug sensitivity is more faithfully characterized at the level of inferred cellular states than through individual molecular markers. Contemporary genome-informed precision oncology relies heavily on the genomic status of individual signaling proteins to guide treatment (Marquart et al., 2018), yet this strategy covers only a fraction of patients. The latent representations produced by ResAttnVAE encode coordinated transcriptional and genomic signatures reflecting underlying pathway activity (Sanchez-Vega et al., 2018), enabling identification of responder subpopulations that cut across traditional cancer type boundaries. Pathway-specific clustering analyses for PI3K/mTOR and EGFR inhibitors corroborate this observation, revealing responder-enriched groups that are not explained by tissue of origin alone.

The attention mechanism adds a further dimension of interpretability by exposing which SGAs most strongly shape the inferred cellular states. Attention highlighted not only canonical oncogenic drivers but also lower-frequency, functionally regulatory genes such as *GNAQ* and *GNA11* (Raamsdonk et al., 2010), suggesting transcriptional influence—rather than mutation prevalence—drives prioritization. Embedding analyses confirmed that genes sharing pathways occupy proximate latent-space regions, echoing pathway-guided architectures such as DrugCell (Kuenzi et al., 2020) and PathDNN (Deng et al., 2020), but without requiring explicit ontology annotations.

Each architectural component contributes a distinct facet of interpretability: residual connections inject SGA context at multiple depths; attention assigns tumor-specific weights, yielding the pathway-coherent prioritization in Fig. 3; and the two-stage design preserves EN coefficient-level transparency—together forming a multi-level framework spanning input, representation, and prediction.

The two-stage design—separating unsupervised representation learning from supervised prediction—confers clear advantages over end-to-end alternatives, consistent with the DREAM challenge finding that modular strategies often outperform monolithic ones (Costello et al., 2014). ResAttnVAE-EN first learns generalizable latent representations via reconstruction, then fits sparse Elastic Net classifiers (Zou and Hastie, 2005), mitigating overfitting while preserving coefficient-level interpretability.

Clinical evaluation in TCGA lung cancer cohorts confirmed that cell-line-trained predictions correlate with both survival and documented treatment response, consistent with prior transfer-learning evidence (Geeleher et al., 2014; Ding et al., 2018). Performance is necessarily constrained by incomplete drug overlap between training and clinical datasets,

but AUROC improved with increasing drug coverage, underscoring the robustness of the approach.

Limitations. Several limitations apply. We did not include external published methods as baselines, since matching drug panels, features, and binarization strategies across studies is non-trivial and left to future work. Attention weights and latent representations afford system-level but correlational, not causal, interpretability. Reliance on bulk expression obscures intratumoral heterogeneity, and evaluation is limited to the GDSC panel (Yang et al., 2013; Iorio et al., 2016), with broader drug libraries or newer modalities (e.g., immunotherapy) untested. Our survival cohort (n=62) is small and retrospective; larger prospective cohorts are needed for definitive validation. Finally, ResAttnVAE-EN requires bulk RNA-seq and mutation/CNV calls, increasingly available in routine profiling; we envision it as a research prototype supporting, not replacing, oncologist decision-making.

In summary, ResAttnVAE-EN advances interpretable machine learning for precision oncology by combining deep generative modeling, attention-based genomic feature weighting, and sparse supervised prediction. The framework provides a biologically grounded and clinically relevant approach to drug sensitivity prediction and offers a flexible foundation for integrating additional molecular and clinical data modalities.

Acknowledgments

We thank Gregory Cooper and Lujia Chen for their support of this work. This study is partially supported by an NIH grant (5R01LM012011). The content is the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

Delora Baptista, Pedro G Ferreira, and Miguel Rocha. Deep learning for drug response prediction in cancer. *Briefings in Bioinformatics*, 22(1):360–379, 2021.

Chunhui Cai, Gregory F Cooper, and Xinghua Lu. Causal discovery of gene regulation with incomplete data. *Journal of the American Medical Informatics Association*, 27(5): 765–775, 2020.

Emerson Y Chen, Vinay Raghunathan, and Vinay Prasad. An overview of cancer drugs approved by the US Food and Drug Administration based on the surrogate end point of response rate. *JAMA Internal Medicine*, 179(7):915–921, 2019.

Yu-Chiao Chiu, Hung-I Harry Chen, Tinghe Zhang, Songjian Zhang, Aparna Gorthi, Li-Ju Wang, Yufei Huang, and Yidong Chen. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Medical Genomics*, 12(Suppl 1):143, 2019.

James C Costello, Laura M Heiser, Elisabeth Georgii, Mehmet Gonen, Michael P Menden, Nicholas J Wang, Mukesh Bansal, Muhammad Ammad ud din, Petteri Hintsanen, Suleiman A Khan, et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nature Biotechnology*, 32(12):1202–1212, 2014.

Lei Deng, Yidong Chen, Yan Zhang, Shuigeng Zhou, et al. Pathway-guided deep neural network toward interpretable and predictive modeling of drug sensitivity. *Journal of Chemical Information and Modeling*, 60(10):4497–4505, 2020.

Ayse Beren Dincer, Safiye Celik, Naozumi Hiranuma, and Su-In Lee. Adversarial deconfounding autoencoder for learning robust gene expression embeddings. *Bioinformatics*, 36(Suppl 2):i573–i582, 2020.

Michael Q Ding, Lujia Chen, Gregory F Cooper, Jonathan D Young, and Xinghua Lu. Precision oncology beyond targeted therapy: combining omics data with machine learning matches the majority of cancer cells to effective therapeutics. *Molecular Cancer Research*, 16(2):269–278, 2018.

Zijian Ding, Songpeng Zu, and Jin Gu. Evaluating the molecule-based prediction of clinical drug responses in cancer. *Bioinformatics*, 32(19):2891–2895, 2016.

Jingcheng Du, Peilin Jia, Yulin Dai, Cui Tao, Zhongming Zhao, and Degui Zhi. Gene2vec: distributed representation of genes based on co-expression. *BMC Genomics*, 20(Suppl 1): 82, 2019.

Luz Garcia-Alonso, Francesco Iorio, Angela Matchan, Nuno Fonseca, Panagiotis Jaaks, Gareth Peat, Miguel Pignatelli, Fiach O'Neill, Julio Saez-Rodriguez, and Mathew J Garnett. Transcription factor activities enhance markers of drug sensitivity in cancer. *Cancer Research*, 78(3):769–780, 2018.

Mathew J Garnett, Elena J Edelman, Sonja J Heidorn, Chris D Greenman, Anahita Dastur, King Wai Lau, Patricia Greninger, Iain Richard Thompson, Xi Luo, Jorge Soares, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, 483(7391):570–575, 2012.

Paul Geeleher, Nancy J Cox, and R Stephanie Huang. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biology*, 15(3):R47, 2014.

Francesco Iorio, Theo A Knijnenburg, Daniel J Vis, Graham R Bignell, Michael P Menden, Michael Schubert, Nanne Aben, Emanuel Goncalves, Syd Barthorpe, Howard Lightfoot, et al. A landscape of pharmacogenomic interactions in cancer. *Cell*, 166(3):740–754, 2016.

Diederik P Kingma and Max Welling. Auto-encoding variational Bayes. In *International Conference on Learning Representations*. ICLR, 2014.

Brent M Kuenzi, Jisoo Park, Samson H Fong, Kyle S Sanchez, John Lee, Jason F Kreisberg, Jianzhu Ma, and Trey Ideker. Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer Cell*, 38(5):672–684, 2020.

Min Li, Yun Wang, Ruiqing Zheng, Xinghua Shi, Yaohang Li, Fang-Xiang Wu, and Jianxin Wang. DeepDSC: A deep learning method to predict drug sensitivity of cancer cell lines. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(2):575–582, 2021.

John Marquart, Emerson Y Chen, and Vinay Prasad. Estimation of the percentage of US patients with cancer who benefit from genome-driven oncology. *JAMA Oncology*, 4(8): 1093–1098, 2018.

Michael P Menden, Francesco Iorio, Mathew Garnett, Ultan McDermott, Cyril H Benes, Pedro J Ballester, and Julio Saez-Rodriguez. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS ONE*, 8(4):e61318, 2013.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, volume 26, pages 3111–3119. Curran Associates, 2013.

Catherine D Van Raamsdonk, Kent G Griewank, Michelle B Crosby, Maria C Garrido, Swapna Vemula, Thomas Wiesner, Anna C Obenauf, Dorothea Wackernagel, Gary Green, Nathalie Bouvier, et al. Mutations in GNA11 in uveal melanoma. *New England Journal of Medicine*, 363(23):2191–2199, 2010.

Shuangxia Ren, Yifeng Tao, Ke Yu, Yifan Xue, Russell Schwartz, and Xinghua Lu. De novo prediction of cell-drug sensitivities using deep learning-based graph regularized matrix factorization. In *Pacific Symposium on Biocomputing*, volume 27, pages 278–289. World Scientific, 2022.

Shuangxia Ren, Gregory F Cooper, Lujia Chen, and Xinghua Lu. An interpretable deep learning framework for genome-informed precision oncology. *Nature Machine Intelligence*, 2024.

Francisco Sanchez-Vega, Marco Mina, Joshua Armenia, Walid K Chatila, Augustin Luna, Konnor C La, Sofia Dimitriadoy, David L Liu, Havish S Kantheti, Sadegh Saghafinia, et al. Oncogenic signaling pathways in The Cancer Genome Atlas. *Cell*, 173(2):321–337, 2018.

Yifeng Tao, Shuangxia Ren, Michael Q Ding, Russell Schwartz, and Xinghua Lu. Predicting drug sensitivity of cancer cell lines via collaborative filtering with contextual attention. In *Machine Learning for Healthcare Conference*, pages 660–684. Proceedings of Machine Learning Research, 2020.

Katarzyna Tomczak, Patrycja Czerwinska, and Maciej Wiznerowicz. The cancer genome atlas (TCGA): an immeasurable source of knowledge. *Contemporary Oncology*, 19(1A): A68–A77, 2015.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pages 5998–6008. Curran Associates, 2017.

Bert Vogelstein, Nickolas Papadopoulos, Victor E Velculescu, Shibin Zhou, Luis A Diaz, and Kenneth W Kinzler. Cancer genome landscapes. *Science*, 339(6127):1546–1558, 2013.

Gregory P Way and Casey S Greene. Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. In *Pacific Symposium on Biocomputing*, volume 23, pages 80–91. World Scientific, 2018.

Yifan Xue, Michael Q Ding, and Xinghua Lu. Learning to encode cellular responses to systematic perturbations with deep generative models. *NPJ Systems Biology and Applications*, 6(1):1–10, 2020.

Wanjuan Yang, Jorge Soares, Patricia Greninger, Elena J Edelman, Howard Lightfoot, Simon Forbes, Nidhi Bindal, Dave Beare, James A Smith, I Richard Thompson, et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research*, 41(D1):D955–D961, 2013.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

Appendix A. Training Hyperparameters

Table 3 reports the training hyperparameters for the representation learning module (ResAttnVAE) and the downstream Elastic Net drug response classifiers, as configured in the training scripts.

Table 3: Training hyperparameters for the ResAttnVAE representation learning module and the downstream Elastic Net drug response classifiers.

Hyperparameter	**Value**
ResAttnVAE representation learning module	
Optimizer	Adam
Learning rate	0.001
Batch size	45
Max epochs	60
Early stopping patience	50 epochs (no improvement in training loss)
Number of attention heads	4
Number of residual/attention blocks	4
Embedding size (hidden blocks)	200
Embedding size (last block)	320
Embedding pruning ratio	0.6 (top-k retained per embedding)
Hidden-layer weight pruning	Enabled, L_1-unstructured, ratio 0.9
Gene2Vec pretraining	Enabled
Loss function	MSE (reconstruction)
Random seed	10
Elastic Net drug response classifiers	
Model	Logistic regression, elastic-net penalty
Solver	SAGA
Regularization strength C and L1_ratio	Selected per drug via grid search with 10-fold CV
Random seed	10

Appendix B. Gene2Vec Embedding Initialization

Gene2Vec is an adaptation of the Word2Vec algorithm (Mikolov et al., 2013) designed for genomic data. Rather than training on text corpora, it learns distributed representations of SGA genes by modeling their co-occurrence patterns across GDSC and TCGA tumor samples. Because SGAs within a single patient lack a defined biological order, random sampling without replacement was performed ten times per sample to construct a training corpus, following the approach of Du et al. (2019). The Word2Vec model from the `gensim.models` module in Python was used with the skip-gram algorithm to learn embeddings. This initialization imposes biologically meaningful structure on the embedding space from the outset and accelerates convergence during training.

Appendix C. Data Preprocessing Details

Somatic Genomic Alterations. Mutation data for GDSC cell lines were obtained from Iorio et al. (2016), while copy number variation (CNV) data and cancer type annotations were acquired from the Cell Model Passports. For TCGA, non-silent somatic mutations were downloaded from the Pan-Cancer Atlas, and CNV estimates were obtained from the Xena portal using the GISTIC2 thresholding method, which categorizes copy number states into five levels ranging from homozygous deletion to high-level amplification. Only homozygous deletions and high-level amplifications (values of ± 2) were considered as CNV events.

Overlap of the curated 1,084-gene panel with the OncoKB Cancer Gene list confirmed substantial concordance across seven independent clinical gene resources, with the highest agreement observed for FoundationOne (307 of 324 genes) and (Vogelstein et al., 2013), 115 of 125 genes).

Gene Expression. Gene expression data for GDSC cell lines consisted of RMA-normalized microarray measurements from Garcia-Alonso et al. (2018), while TCGA expression profiles were derived from RNA sequencing data downloaded from the Xena portal. Genes were selected following the procedure described by Ding et al. (2018), which identifies high-variance genes through a combination of Hartigans' dip test for multimodal distributions, the outlier sum method for unimodal distributions with notable outlier populations, and median absolute deviation for genes exhibiting high variance regardless of distribution shape.

Because microarray and RNA-seq platforms yield systematically different distributions, a nonparanormal (NPN) transformation was applied independently to each dataset, converting marginal distributions to approximate normality while preserving rank-based relationships. Min–max normalization was then applied to rescale all values to the $[0, 1]$ interval.

Batch Effects. Within GDSC, expression and SGA data come from the same research center, so cross-sample batch effects are not a major concern. For GDSC-to-TCGA transfer, the NPN transformation maps each gene's marginal distribution to a common normal scale independently within each dataset, removing platform-level shifts while preserving rank-based relationships, without requiring shared calibration samples. Residual platform-specific variation may nonetheless remain, which is difficult to rule out without shared reference samples.

Drug Sensitivity Binarization. Drug sensitivity in GDSC was summarized using activity area (AA). The GDSC1 dataset contains 367 compounds; drugs sharing the same name but different identifiers were treated as distinct entities. Responses were discretized into sensitive and resistant categories using the waterfall method (Ding et al., 2018). Specifically, sensitivity measurements across all cell lines for a given drug are sorted to produce a waterfall distribution, a linear regression is fitted, and the goodness of fit is evaluated by Pearson correlation. If the correlation falls below 0.95, the major inflection point—defined as the point of maximal distance from a line connecting the distribution endpoints—serves as the cutoff; otherwise, the median value is used as the threshold.

Appendix D. Clinical Prediction Details

Survival Analysis. Clinical data including treatments and overall survival were collected for 62 LUAD and LUSC patients from the TCGA consortium who received different combinations of cisplatin (41 cases), pemetrexed (19 cases), paclitaxel (17 cases), and vinorelbine (10 cases) as adjuvant therapies (Ren et al., 2022). EN models trained on GDSC cell line data were applied to each patient to derive probabilities of sensitivity to the drugs in their prescribed regimen. Because predicted probabilities across different drugs were not well calibrated, a patient was designated as sensitive to a given drug if the prediction probability fell within the top 40th percentile of all patients treated with that drug, consistent with the approximate 40% average response rate reported for FDA-approved cancer drugs (Chen et al., 2019). A patient was then classified as a responder if predicted sensitive to any drug in the regimen, and as a non-responder otherwise.

Drug Response Prediction in TCGA. Curated TCGA drug response data were obtained from Ding et al. (2016). From the original dataset of 2,572 patient–drug response pairs, records with complete SGA data were retained and analysis was restricted to 27 drugs shared between TCGA and GDSC. After removing records for drugs absent from GDSC, eliminating duplicates, and excluding patients with inconsistent responses, the final dataset comprised 1,133 records across 762 patients. Clinical responses were dichotomized into responders (complete or partial response) and non-responders (stable or progressive disease) (Geeleher et al., 2014).

For each patient–drug combination, the model predicted a drug sensitivity probability. For patients receiving multiple drugs, overall sensitivity was computed using a noisy-or aggregation:

$$P(y = 1) = 1 - \prod_k (1 - P(y = 1|Drug_k)) \tag{10}$$

where the product runs over all drugs with available predictions. Drugs not represented in the GDSC training set were omitted by assuming $P(y = 1|Drug_{omit}) = 0$. A drug coverage ratio was defined for each patient as the fraction of administered drugs for which predictions were available. AUROC values were computed across varying drug coverage ratio thresholds.

Appendix E. Embedding Connectivity Map Construction

For the SGA gene connectivity map, the two embedding vectors for each SGA gene were concatenated into a single representation, and pairwise cosine similarities were computed. For each gene, the ten most similar neighbors were identified, and edges were added between gene pairs if one or both ranked among the other's nearest neighbors. Arrows indicate directionality of the neighbor relationship, with double-headed arrows denoting mutual nearest-neighbor status. Edge thickness corresponds to cosine similarity strength.

For the drug connectivity map, the Elastic Net coefficient vectors were treated as drug-level embeddings reflecting the relative importance of features for predicting sensitivity to each compound. Pairwise cosine similarities were computed, and edges were added between drug pairs ranking among each other's top five nearest neighbors.