

A causal inference framework for identifying essential genes to enhance drug synergy prediction

Huaiwu Zhang^{1, }, Xinliang Sun², Jianxin Wang^{2, }, Min Li^{2, }, Jing Tang^{1,*, }

¹Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki, 00290, Finland

²School of Computer Science and Engineering, Central South University, Changsha, 410083, China

*Corresponding authors. Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Haartmaninkatu 8, Helsinki, 00290, Finland.

E-mail: jing.tang@helsinki.fi

Associate Editor: Jianlin Cheng

Abstract

Motivation: Identifying synergistic drug combinations holds promise for more effective treatment strategies. Recent deep learning methods such as Transformers and Graph Neural Networks have shown improved predictive performance, but most of them integrate drug and cell line representations without explicitly modelling the causal effects of genes in mediating drug responses.

Results: We introduce CADS (Causal Adjustment for Drug Synergy), a deep learning framework that explicitly models the gene–drug causal relationships to improve both prediction accuracy and biological interpretability. CADS integrates multi-omics data with a learnable gene-selection mechanism that performs causal backdoor adjustment, enabling both drug synergy prediction and causal gene discovery. Across multiple benchmark datasets, CADS consistently achieves superior performance compared with state-of-the-art drug synergy prediction models. In addition, downstream analyses on case studies demonstrate that the inferred gene causal scores can recover clinically validated cancer-related genes involved in drug combinations. These results demonstrate that explicitly modelling causal genetic effects can enhance the reliability and interpretability of drug synergy prediction.

Availability and implementation: The source code of CADS can be found at <https://github.com/HuaiwuZhang/causalDC>.

1 Introduction

Combination therapies have become a promising strategy for the treatment of complex diseases such as cancer and infectious diseases (Al-Lazikani *et al.* 2012). By simultaneously perturbing multiple biological pathways, drug combinations can improve therapeutic efficacy and reduce the emergence of drug resistance compared to monotherapies (Gilad *et al.* 2021). Despite these advantages, identifying truly synergistic drug pairs remains challenging, as most clinically-approved combinations exhibit additive or independent effects rather than biologically meaningful interactions (Bulusu *et al.* 2016, Plana *et al.* 2022).

Experimental assay of drug combinations is commonly performed using high-throughput screening (HTS) platforms. Although effective on a small scale, HTS becomes increasingly infeasible as the combinatorial search space grows exponentially (Sadybekov and Katritch 2023). Furthermore, HTS screens often provide limited insight into the molecular mechanisms underlying drug interactions, thereby restricting their utility for rational drug design and clinical translation (Broach and Throner 1996). These limitations have motivated the development of

computational approaches that can efficiently prioritize candidate drug combinations while offering mechanistic interpretations, thus establishing causal links with biological pathways and biomarkers.

With the increasing availability of large-scale pharmacogenomics resources such as DrugComb (Zagidullin *et al.* 2019), computational prediction of drug synergy has advanced rapidly. Early studies often relied on classical machine learning models, including random forests (Jeon *et al.* 2018), support vector machines (SVMs) (Noble 2006), and gradient-boosting methods (Chen and Guestrin 2016), which integrated features such as drug chemical descriptors and genomic profiles of cell lines (Wu *et al.* 2022). For instance, ERT (Jeon *et al.* 2018) integrated monotherapy responses and synthetic lethality data with random forests and SVMs to predict anticancer synergy, while DCPT (He *et al.* 2018) leveraged XGBoost to model selective synergy based on compound sensitivity differences. While these approaches demonstrated improved results, their performance often degraded when applied to unseen drug combinations or untested biological contexts, reflecting limited generalization capability (Partin *et al.* 2023).

Received: 9 October 2025. Revised: 17 December 2025. Accepted: 6 January 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

More recently, deep learning methods have emerged as the dominant paradigm for drug synergy prediction, achieving improved predictive accuracy across multiple benchmarks (Ahmed et al. 2023). Representative methods such as DeepSynergy (Preuer et al. 2018) and MatchMaker (Kuru et al. 2022) introduced multi-branch neural architectures that jointly encode drug fingerprints and gene expression profiles. Subsequent work expanded this framework by incorporating drug-induced transcriptomic responses (EC-DFR (Lin et al. 2022)), graph neural networks for molecular structure modeling (DeepDDS (Wang et al. 2022)), hypergraph representations of drug-cell interactions (HypergraphSynergy (Liu et al. 2022)), and biological prior knowledge such as drug-target and protein-protein interaction networks (Jiang et al. 2020; Yang et al. 2021; Yu et al. 2022).

To further enhance the prediction performance, recent studies have focused on cross-modal interaction modeling, using attention mechanisms and Transformer-based architectures to capture complex dependencies between drugs, genes, and cellular contexts (Hu et al. 2022, Xu et al. 2023, Rafiei et al. 2023). In particular, SynergyX incorporates dynamic cross-modal interaction modeling together with biological network information (Guo et al. 2024), while PermuteDDS adopts a complementary strategy by constructing rich, multi-dimensional drug representations from diverse fingerprint types (Zhao et al. 2024). More recently, DVHSyn learns drug-drug-cell triplet context on a hypergraph and its expanded heterogeneous graph with selective fusion (Yu et al. 2026), while MDI-DDI integrates both the 2D and 3D structural information of drugs (Gong et al. 2025). In addition, MOSAIC splits SMILES into meaningful fragments and performs multi-granularity prediction (Zhang et al. 2025), and BAITSAO (Liu et al. 2025) uses LLM-derived drug and cell-line embeddings together with multi-task pre-training on large synergy datasets. Despite these methodological advances, two fundamental challenges remain unresolved. First, many models do not explicitly distinguish causal genetic factors from spurious correlations introduced by the high-dimensionality of cell line molecular features, leading to unstable predictions across datasets. Second, most existing approaches focus primarily on predictive performance, offering limited interpretability about which genes causally affect drug synergy.

In this study, we proposed CADS, a deep learning framework to explicitly model the causal effects of genes in drug synergy. CADS integrates multiple drug representations, including SMILES strings and chemical fingerprints, with multi-omics cell line profiles through a novel multi-modal fusion strategy. By constructing a structural causal model and applying backdoor adjustment via a learnable masking mechanism, CADS separates causal genes from confounding genetic factors during representation learning. This design enables the model to achieve higher predictive accuracy while simultaneously identifying genes that are causally relevant to drug interactions. A preliminary version of this study has appeared in the ISBRA 2025 conference (Zhang and Tang 2025). Compared to the conference version, we have made major extensions including:

- We incorporated CADS into multiple baselines and conducted a more comprehensive evaluation of predictive performance and robustness.

- We performed an interpretability analysis which revealed that embedding CADS strengthens cell line data representations to improve performance across various metrics.
- We highlighted the capability of CADS to uncover novel drug combinations and drug targets through substantial case studies.

The paper is organized as follows. In Section 2 we began with causal inference by constructing a *structural causal model* (SCM) (Figure 1, available as [supplementary data](#) at *Bioinformatics* online), and then designed a deep learning framework for the prediction of drug synergy. In Section 3 we integrated CADS into several carefully selected baseline models and demonstrated its predictive performance through comprehensive experiments. We then conducted ablation and few-shot analyses to validate the CADS's performance and robustness, together with multiple case studies to identify and interpret causal genes associated with drug synergy. Section 4 provides a detailed discussion of the results and the limitations of the proposed approach.

2 Methodology

In this section, we presented the details of our CADS framework. We introduced CADS, a novel deep learning module designed to explicitly model causal relationships in drug-cell interactions. This component is systematically embedded into three state-of-the-art baseline frameworks including: DeepSynergy (Preuer et al. 2018), DeepDDS (Wang et al. 2022), and SynergyX (Guo et al. 2024).

2.1 Framework overview

Figure 1 shows the framework of CADS. Previous methods for drug synergy prediction typically employ SMILES sequence (Guo et al. 2024) or 2D graph (Wang et al. 2022) to represent global features of drugs, followed by feature concatenation with cell line characteristics or feature fusion using a cross-attention mechanism. It is noteworthy that in drug synergy prediction, when specific drug combinations are regarded as the primary entity, the intercellular contextual information C corresponding to each sample is incorporated into the encoding process as auxiliary information. To enhance conciseness of notation, we denoted this encoding process as $\mathcal{F}$ throughout this paper.

As illustrated in Appendix Section C, available as [supplementary data](#) at *Bioinformatics* online, the sole extraction of cell line information may introduce potential confounding factors. Therefore, when modeling drug combinations and cell line representations, we adopted a causal decoupling strategy: decomposing the encoder $\mathcal{F}$ into two parallel branches—a causal encoder $\mathcal{F}_c$ and a trivial encoder $\mathcal{F}_t$, which maintain identical architectures but possess non-shared parameters. This process can be formally represented as follows:

$$H_c = \mathcal{F}_c(D_A, D_B; G_c), \quad H_t = \mathcal{F}_t(D_A, D_B; G_t), \quad (1)$$

Where G_c and G_t represent causal cell information and trivial cell information. The embeddings H_c and H_t represent the final

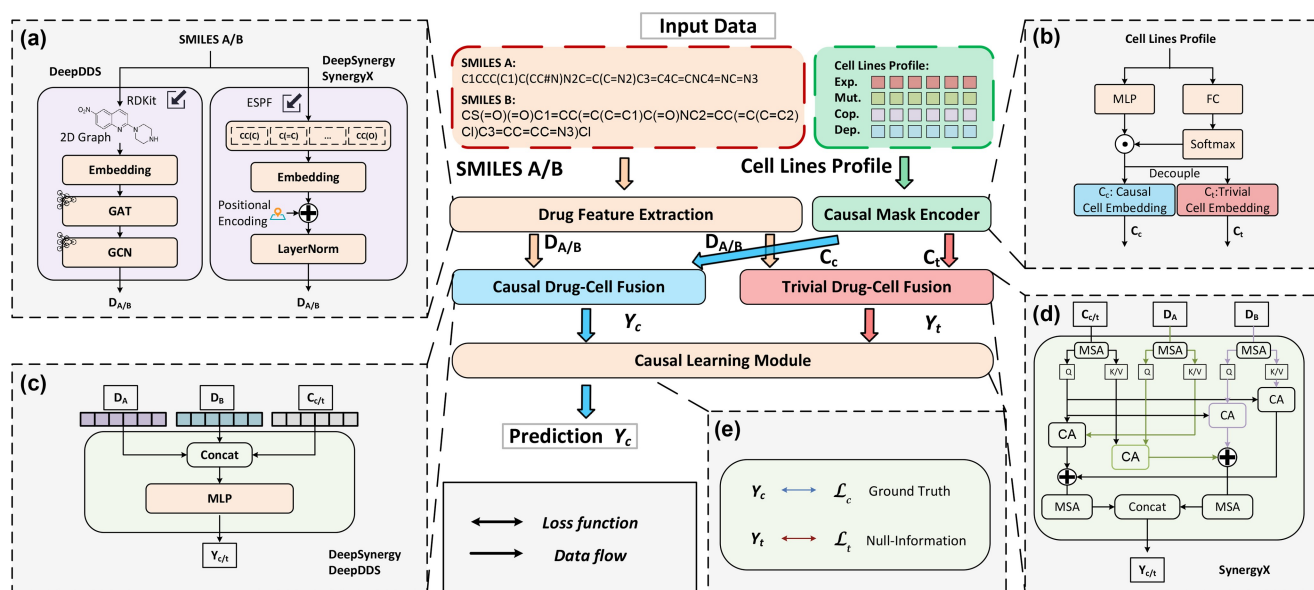


Figure 1 The pipeline of CADS. The CADS framework primarily consists of three main components: (a) *Drug Feature Extraction* is employed to extract drug features from sequences or 2D molecular graphs; (b) *Causal Mask Encoder* generates learnable causal mask pairs to decouple causal information from trivial information leveraging cell line profiles; (c–d) *Drug-Cell Fusion* employs Multi-Layer Perceptrons (MLP) or Cross-Attention (CA) and Multi-head Self-Attention (MSA) to fuse drug features with decoupled cell representations; (e) *Causal Learning Module* learns hard-to-observe causal information by training the causal and trivial components separately with ground truth and null-information labels, respectively.

encoded representations of drug pairs after integrating causal information and trivial information, respectively.

Next, under the guidance of a specifically designed *Causal Mask Mechanism*, the output features of the two encoders are disentangled through separate processing pathways. H_c and H_t are fed into dual processing paths:

- *Causal branch*: Extracts invariant representations (causal features) from H_c .
- *Trivial branch*: Captures confounding effects (trivial features) from H_t .

Finally, a *Causal Learning Module* subsequently performs representation-level interventions to coordinate with downstream tasks, enabling explicit separation between causal features and confounding factors to achieve robust representation learning. We further explained and illustrated the changes to the model architecture before and after CADSEmbedding in Appendix Section D and Fig. 2, available as [supplementary data](#) at *Bioinformatics* online to make them easier to visualize.

In the subsequent sections, we presented comprehensive details on the design of key modules and their integration methodology into the baseline framework.

2.2 Drug feature extraction

As shown in Fig. 1a, this module performs feature extraction on drug SMILES sequences. Due to the distinct drug representation handling mechanisms across the selected baseline models (Sequence or 2D-graph), we implemented two separate sets of feature extraction methodologies following prior research. As illustrated in Appendix Section E.2.1, available as [supplementary data](#) at *Bioinformatics* online, for SynergyX and DeepSynergy

which process sequential structures, we segmented extended SMILES sequences into interpretable substructures using tokenization methodologies analogous to those in natural language processing (Huang et al. 2019), with each substructure treated as a discrete token ($\mathbf{T}$):

$$\mathbf{D}_k = \mathcal{N}(\mathbf{T} \oplus \mathbf{P}), \quad k \in \mathcal{K} := \{A, B\}, \quad (2)$$

Where $\mathbf{D}_k$ is the representation of Drug A or Drug B, $\mathcal{N}$ means layer normalization, $\mathbf{P}$ represents the positional encoding matrix satisfying $\|\mathbf{P}_i\|_2 = 1$.

For DeepDDS processing 2D molecular graph structures, SMILES representations were converted into graph topologies using the RDKit library, followed by cascaded GAT (A) (Velićković et al. 2017) and GCN (G) (Kipf and Welling, 2018) architectures to capture local and global graph features respectively:

$$\mathbf{D}_k = \mathcal{G}^{\circ} \mathcal{A}(\mathbf{V}^{(k)}, \mathbf{E}^{(k)}), \quad k \in \mathcal{K} := \{A, B\} \quad (3)$$

Where $\mathbf{V}^{(k)}$ and $\mathbf{E}^{(k)}$ represents denote the set of atoms and chemical bonds, respectively.

2.3 Causal mask encoder

Concurrently with drug feature extraction, feature encoding of cell lines was performed in parallel. Appendix Section C.1, available as [supplementary data](#) at *Bioinformatics* online highlights that filtering out trivial genes is essential for uncovering true causal impacts in the context of drug synergy. Nevertheless, explicitly distinguishing between causal and trivial genes remains a significant challenge in real-world scenarios. To overcome this

limitation, we developed a computational framework designed to refine parametric masks via discriminative learning. Specifically, we employed a *causal awareness mask* ($\mathbf{M}_c$) alongside a *trivial awareness mask* ($\mathbf{M}_t$) to direct the distinct decoupling of features throughout parallel branches. For any given cell line profile $\mathbf{C} \in \mathbb{R}^{F \times N}$, a soft mask $\mathbf{M} \in \mathbb{R}^{1 \times N}$ was allocated. Within this mask, individual elements denote the relative significance of their respective genes, generally scaled to a range of 0–1. Given a mask $\mathbf{M}$, its complementary mask is defined as

$$\bar{\mathbf{M}} = \mathbf{1}_N - \mathbf{M}, \quad (4)$$

Where $\mathbf{1}_N$ denotes an all-ones matrix. In formal terms, the collection of all causal genes constitutes the causal gene set, whereas the remaining involved genes are grouped into the trivial gene set. To learn these masks through the cell lines information while ensuring traceability to individual genes, we first performed feature extraction from cell lines using a fully-connected neural network. For each gene g_i , we computed its attention scores as follows:

$$\alpha_i = \left[\frac{e^{(\mathbf{W}\mathbf{g}_i)_1}}{e^{(\mathbf{W}\mathbf{g}_i)_1} + e^{(\mathbf{W}\mathbf{g}_i)_2}} \right] = \begin{bmatrix} \alpha_{c_i} \\ \alpha_{t_i} \end{bmatrix} \quad (5)$$

for $i = 1, \dots, N$

Where α_{c_i} and α_{t_i} denote the attention weights at the gene level for a specific gene g_i , corresponding to the causal and trivial masks, respectively. We apply a softmax across the two-channel vector for each gene i , yielding gene-level attention scores α_{c_i} and α_{t_i} for the causal and trivial masks, respectively. By construction, this ensures $\alpha_{c_i} + \alpha_{t_i} = 1$. These attention scores indicate the model's attention to each gene within the corresponding cell-line feature matrix. We then construct a set of soft masks based on these scores:

$$\mathbf{M}_c = [\alpha_{c_i}]_{i=1}^N, \quad \mathbf{M}_t = [\alpha_{t_i}]_{i=1}^N, \quad \mathbf{M}_c + \mathbf{M}_t = \mathbf{1}_N \quad (6)$$

Therefore, our primary goal was to optimize the learnable representations $\mathbf{G}_c = \mathbf{C} \odot \mathbf{M}_c$ and $\mathbf{G}_t = \mathbf{C} \odot \mathbf{M}_t$ for the effective separation of causal and trivial genes, with $\odot$ indicating the Hadamard product.

Up to now, we have captured causal features within cell line representations by allocating gene-level attention scores. Besides drug and cell representations ($\mathbf{D}_{A/B}$; $\mathbf{G}_{c/t}$) have been mapped into a shared feature space through the projection process.

2.4 Drug-cell fusion

In this section, we introduced a dual-branch parallel architecture (Causal/Trivial Drug-Cell Fusion) to implement cross-modal alignment between pharmacological data and cellular gene profiles, thereby connecting drug features with causal/trivial gene modalities respectively through inter-modal interactions.

As shown in Fig. 1d, for SynergyX, we further employ a set of cross-attention (CA) and multihead self-attention (MSA) mechanisms:

$$\begin{aligned} \mathcal{CA}(\mathbf{X}_1, \mathbf{X}_2) &= \sigma \left(\frac{(\mathbf{X}_1 \mathbf{W}_h^Q)(\mathbf{X}_2 \mathbf{W}_h^K)^\top}{\sqrt{d_k}} \right) (\mathbf{X}_2 \mathbf{W}_h^V), \\ \mathcal{MSA}(\mathbf{X}) &= \sigma \left(\frac{(\mathbf{X} \mathbf{W}_h^Q)(\mathbf{X} \mathbf{W}_h^K)^\top}{\sqrt{d_k}} \right) (\mathbf{X} \mathbf{W}_h^V), \end{aligned} \quad (7)$$

Where $\sigma(\cdot)$ denotes the softmax function. First, cross-modal communication is executed between the drug and cell line representations:

$$\begin{aligned} \mathbf{G}'_i &= \mathcal{CA}(\mathbf{G}_i, \mathbf{D}_k), \\ \mathbf{D}'_k &= \mathcal{CA}(\mathbf{D}_k, \mathbf{G}_i); \\ \mathbf{G}_i &= \mathbf{G}'_i \\ i &\in \{c, t\}, k \in \{A, B\} \end{aligned} \quad (8)$$

Furthermore, recognizing that drug synergy is deeply governed by underlying pharmacological and pharmacokinetic mechanisms, we systematically model the interactive features between the two drugs, which is a crucial step:

$$\mathbf{D}_i = \mathcal{CA}(\mathbf{D}'_i, \mathbf{D}'_k); i, k \in \mathcal{K} := \{A, B\}, i \neq k \quad (9)$$

For $i, k \in \mathcal{K} = A, B$ with $i \neq k$, we apply cross-attention in both directions: $\mathbf{D}'_A = \mathcal{CA}(\mathbf{D}'_A, \mathbf{D}'_B)$ and $\mathbf{D}'_B = \mathcal{CA}(\mathbf{D}'_B, \mathbf{D}'_A)$; thus $\mathbf{D}'_A$ attends to $\mathbf{D}'_B$ and vice versa, and both directions are used. Next, for all three baseline models, we adhered to the original methodology by concatenating ($\parallel$) $\mathbf{D}_{A/B}$ with $\mathbf{G}_{c/t}$, followed by MLP-based prediction $\mathcal{Y}_i$ in our implementation:

$$\mathcal{Y}_i = \varphi^\circ \text{MLP}(\mathbf{G}_i \parallel \mathbf{D}_A \parallel \mathbf{D}_B), \quad i \in \{c, t\} \quad (10)$$

Where φ represents the $\text{Tanh}(\cdot)$ activation functions.

2.5 Causal learning module

Through the previously established dual-branch parallel architecture, we derived corresponding drug synergy score predicted based on Causal and Trivial genes, respectively ($\mathcal{Y}_i$, $\mathcal{Y}_i$). Considering that it is difficult to manually identify causal versus trivial genes discussed in Appendix Section C.2, available as [supplementary data](#) at *Bioinformatics* online, we leveraged the causal masking mechanism described in Section 2.3 to conduct joint training, thereby achieving gene disentanglement between Causal and Trivial genetic components.

As established by the Structural Causal Model (SCM), causal genes exert genuine influence on drug synergy. Consequently, the output of the causal branch undergoes regression training using ground-truth labels ($\hat{\mathcal{Y}}_c$). In contrast, trivial genes are designed to encapsulate mechanisms orthogonal to drug synergy. Therefore, the trivial branch is trained with all-zero values as null-information labels ($\hat{\mathcal{Y}}_t$), with the objective of rendering this branch incapable of discriminating between synergistic and antagonistic effects. Consequently, we defined the loss functions as follows:

$$\mathcal{L} = \mathcal{H}(\mathcal{Y}_c, \hat{\mathcal{Y}}_c) + \mathcal{H}(\mathcal{Y}_t, \hat{\mathcal{Y}}_t), \hat{\mathcal{Y}}_t = \mathbf{0}, \quad (11)$$

With $\mathcal{H}(\cdot, \cdot)$ denoting the standard Mean Squared Error (MSE) loss function. By leveraging backpropagation, we imposed stability conditions on the causal genes and independence requirements on the trivial ones. This strategy effectively mitigates spurious correlations, thereby yielding more reliable drug synergy predictions.

3 Experiments & results

3.1 Overall performance

Table 1 presents the overall performance across five distinct scores (See also Appendix [Section E.2.2](#), available as [supplementary data](#) at *Bioinformatics* online) under varying model configurations. Specifically, we reported the performance results for the original baseline model and the model employing the CADS framework (+CADS). The experimental results reveal a clear overarching trend: performance increased across all three baseline models following CADS integration, with a more pronounced improvement observed in the architecturally more complex SynergyX model. Notably, in the predictive performance assessment using multiple scores (*i.e.*, Loewe, HSA), the simpler DeepSynergy architecture outperformed the more complex SynergyX prior to CADS embedding. However, post-CADS integration, the performance of the complex SynergyX model surpassed that of the simpler model. A critical observation emerges when comparing how the two backbone models benefit from CADS. While DepSynergy+CADS showed performance gains, these improvements were statistically significant in only a subset of the evaluation metrics (specifically Loewe and ZIP). In contrast, the integration of CADS into SynergyX resulted in statistically significant improvements ($p < 0.05$) across a much broader array of indicators simultaneously. This phenomenon highlights a key interaction between model complexity and causal constraints. SynergyX, with its Transformer-based architecture, possesses a significantly larger parameter space and higher model capacity than the MLP-based DeepSynergy. While this high capacity allows SynergyX to capture intricate molecular patterns, it also renders the model more susceptible to learning spurious correlations, which can lead to inconsistent performance across different evaluation dimensions. This suggests that complex models may be prone to overfitting cell line contexts, potentially introducing additional confounding factors. In contrast, embedding CADS effectively disentangles these confounding factors, thereby enabling the complex model to realize its superior performance potential. As deep learning models for drug synergy become more complex, causal alignment becomes increasingly vital to ensure comprehensive and reliable performance.

3.2 Causal genes ablation

To validate whether the model relies more heavily on causal genes during prediction, we conducted ablation experiments targeting both causal genes and trivial genes. We performed testing across five independent replicates of the CADS

embedded within SynergyX (embedding five synergy metrics) by applying zero-masking to the top 1000 genes within the designated causal group and trivial group for each drug combination-cell line pair. As shown in [Fig. 2](#), our analysis revealed that model performance is superior following ablation of the top trivial genes compared to ablation of the top causal genes. This finding indicates that masking causal gene information resulted in greater overall performance degradation, demonstrating that the model predominantly depends on causal gene information when generating predictions.

3.3 Robustness test

In this section, to analyse the robustness of CADS, we introduced few-shot learning and provided a demonstration of its application potential for novel drug combination discovery.

3.3.1 Few-shot learning

To validate whether CADS can accurately infer cell-drug interactions under limited data scenarios, we employed a few-shot learning strategy, training the model using selected datasets [0.3, 0.5, 0.7] according to the method described in Appendix [Section E.2.3](#), available as [supplementary data](#) at *Bioinformatics* online.

To statistically quantify the benefit of the causal mechanism under data-scarce conditions, we conducted a linear regression analysis on the performance curves relative to the training set size. As illustrated in [Fig. 3](#), we fitted the performance trends of both the baseline and the CADS-integrated models using linear functions ($y = kx + b$). The slope k serves as a critical indicator of the model's sensitivity to data availability. A steeper slope implies a rapid degradation in performance as data decreases. Our analysis reveals that the absolute value of the slope for the CADS-integrated model ($|k_{CADS}|$) is notably smaller than that of the baseline ($|k_{Base}|$). This “flattened” trend line demonstrates that the CADS module effectively enhances the model's robustness. It attenuates the rate of performance degradation when training samples are limited, thereby validating that the causal mechanism plays a vital role in maintaining predictive stability in few-shot learning tasks. [Table 4](#), available as [supplementary data](#) at *Bioinformatics* online shows the few-shot learning results of the three baselines with CADS embedded.

3.3.2 Novel drug combination discovery

To further investigate the robustness of CADS and its generalization capability for novel drug combinations, we assessed the capacity of CADS for both novel drug combination discovery and drug repurposing. Given the vast unexplored drug space, we selected several commonly used cancer cell lines, including HCT116 (Colon cancer), A549 (Lung cancer) and MDA-MB-231 (Breast cancer) according to the method described in Appendix [Section E.2.4](#), available as [supplementary data](#) at *Bioinformatics* online. For drug combinations, we chose those present in the dataset but without documented experimental data paired with these cell lines. Through the aforementioned screening methodology, we identified a total of 20,338 novel drug combinations across the three cell lines. We utilized CADS trained on five independent replicated trials to predict the synergy of the aforementioned novel drug combinations. [Table 2](#) presents the top four

Table 1 Performance comparison of the three baseline models after embedding CADs across five tasks with five runs.

Score	Model	Validating				Testing			
		MAE	RMSE	R ²	PCC	MAE	RMSE	R ²	PCC
Loewe	DeepSynergy	6.236± 0.079	9.097± 0.095	0.624± 0.008	0.792± 0.005	6.224± 0.079	9.063± 0.100	0.623± 0.008	0.792± 0.006
	+ CADs	6.185 ± 0.035	9.047 ± 0.047	0.628 ± 0.004	0.796 ± 0.004	6.155 ± 0.033	8.997 ± 0.039	0.628 ± 0.003	0.796 ± 0.003
	DeepDDS	6.268± 0.056	9.130± 0.064	0.621± 0.005	0.791± 0.004	6.269± 0.063	9.156± 0.069	0.615± 0.006	0.787± 0.005
	+ CADs	6.200± 0.039	9.048± 0.039	0.628 ± 0.003	0.796 ± 0.002	6.196± 0.046	9.060± 0.050	0.623± 0.004	0.794± 0.002
	SynergyX	6.288± 0.115	9.240± 0.096	0.612± 0.008	0.787± 0.004	6.238± 0.106	9.195± 0.094	0.612± 0.008	0.787± 0.004
	+ CADs	5.978 0.069	8.900 0.077	0.640 0.006	0.801 0.004	5.946 0.082	8.855 0.087	0.640 0.007	0.800 0.004
HSA	DeepSynergy	3.998± 0.072	5.948± 0.086	0.787± 0.006	0.888± 0.004	3.955± 0.073	5.893± 0.085	0.790± 0.006	0.889± 0.003
	+ CADs	3.963 ± 0.065	5.907 ± 0.064	0.790 ± 0.005	0.889 ± 0.003	3.931± 0.072	5.867 ± 0.074	0.792 ± 0.005	0.890 ± 0.003
	DeepDDS	4.080± 0.027	6.059± 0.021	0.779± 0.002	0.883± 0.001	4.060± 0.029	6.050± 0.022	0.779± 0.002	0.883± 0.001
	+ CADs	3.985± 0.080	5.972± 0.079	0.786± 0.006	0.887± 0.003	3.971± 0.076	5.971± 0.076	0.785± 0.005	0.886± 0.003
	SynergyX	3.965 ± 0.149	6.058± 0.165	0.779± 0.012	0.887± 0.004	3.928± 0.136	6.029± 0.158	0.781± 0.011	0.888± 0.004
	+ CADs	3.755 0.021	5.806 0.026	0.797 0.002	0.893 0.001	3.731 0.028	5.771 0.038	0.799 0.003	0.894 0.001
Bliss	DeepSynergy	4.540± 0.137	6.949± 0.179	0.880± 0.006	0.938± 0.003	4.538± 0.139	7.059± 0.183	0.877± 0.006	0.937± 0.003
	+ CADs	4.466± 0.086	6.861± 0.102	0.883± 0.003	0.940± 0.002	4.472± 0.078	6.977± 0.098	0.880± 0.003	0.939± 0.002
	DeepDDS	4.698± 0.094	e7.116± 0.148	0.874± 0.005	0.935± 0.003	4.720± 0.104	7.269± 0.128	0.870± 0.005	0.933± 0.002
	+ CADs	4.480± 0.069	6.850± 0.069	0.883± 0.002	0.940± 0.001	4.491± 0.076	7.024± 0.094	0.879± 0.003	0.938± 0.002
	SynergyX	4.324 ± 0.140	6.786 ± 0.153	0.885 ± 0.005	0.942 ± 0.002	4.311 ± 0.138	6.932 ± 0.159	0.882 ± 0.005	0.940 ± 0.002
	+ CADs	4.134 0.029	6.588 0.039	0.892 0.001	0.944 0.001	4.144 0.021	6.745 0.014	0.888 0.000	0.942 0.000
ZIP	DeepSynergy	3.434± 0.057	5.244± 0.061	0.782± 0.005	0.885± 0.003	3.438± 0.059	5.380± 0.068	0.776± 0.006	0.881± 0.003
	+ CADs	3.371 ± 0.012	5.174 0.014	0.788 0.001	0.888 0.001	3.374 ± 0.013	5.313± 0.019	0.782± 0.002	0.885± 0.001
	DeepDDS	3.576± 0.068	5.402± 0.076	0.769± 0.007	0.877± 0.004	3.585± 0.081	5.556± 0.087	0.762± 0.007	0.873± 0.004
	+ CADs	3.423± 0.041	5.253± 0.044	0.781± 0.004	0.884± 0.002	3.436± 0.049	5.444± 0.034	0.771± 0.003	0.879± 0.002
	SynergyX	3.409± 0.029	5.277± 0.041	0.779± 0.003	0.886± 0.001	3.411± 0.022	5.440± 0.028	0.771± 0.002	0.881± 0.001
	+ CADs	3.339 0.016	5.193 ± 0.018	0.786 ± 0.002	0.887 ± 0.001	3.338 0.008	5.354 ± 0.009	0.779 ± 0.001	0.883 ± 0.001
S-Score	DeepSynergy	6.134± 0.103	8.442 ± 0.111	0.761 ± 0.006	0.873± 0.004	6.187± 0.118	8.544± 0.126	0.758± 0.007	0.872± 0.004
	+ CADs	6.058± 0.037	8.375± 0.050	0.765± 0.003	0.875± 0.002	6.095± 0.044	8.452± 0.053	0.763± 0.003	0.874± 0.002
	DeepDDS	6.129± 0.034	8.468± 0.024	0.760± 0.001	0.873± 0.001	6.180± 0.026	8.563± 0.030	0.757± 0.002	0.871± 0.001
	+ CADs	6.107± 0.021	8.443± 0.033	0.761 ± 0.002	0.874 ± 0.001	6.144 ± 0.028	8.528 ± 0.034	0.759 ± 0.002	0.873 ± 0.001
	SynergyX	6.290± 0.037	8.710± 0.039	0.746± 0.002	0.868± 0.001	6.325± 0.028	8.801± 0.041	0.743± 0.002	0.867± 0.001
	+ CADs	6.086 ± 0.078	8.502± 0.089	0.758± 0.005	0.871± 0.002	6.158± 0.076	8.634± 0.084	0.753± 0.005	0.869± 0.002

Bold and underlined values indicate the best and second-best results in specific task, respectively. The colors **RED** and **PINK** denote statistically significant performance gains relative to the post-CADS embedding state, corresponding to t-test (Kim 2015) *P*-values below 0.01 and 0.05, respectively.

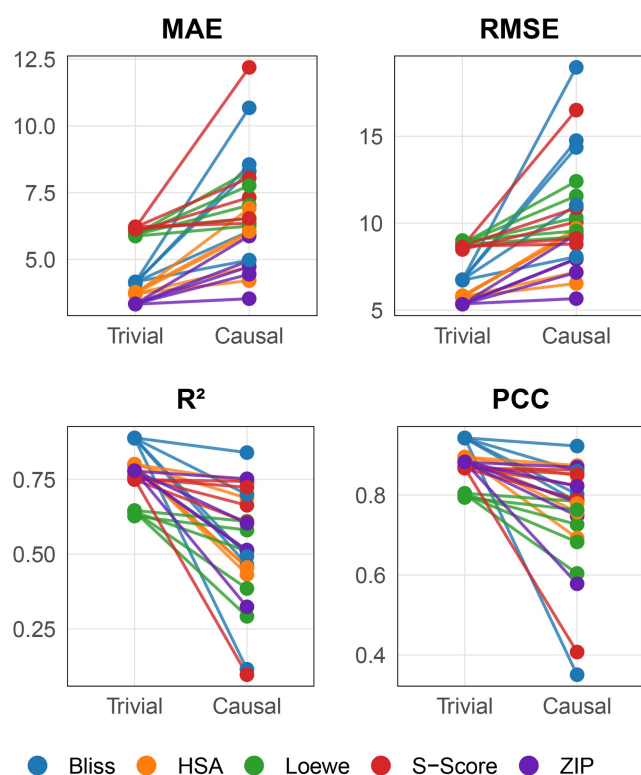


Figure 2 The Paired Comparison Plot illustrates the performance losses in four evaluation metrics (MAE, RMSE, R^2 , PCC) for five synergy metrics (Bliss, HSA, Loewe, S-Score, ZIP) following the zero-masking of the top 1000 trivial genes and the top 1000 causal genes.

novel drug combinations, ranked by predicted S-Scores, across three distinct cell lines.

To validate these predictions, we performed an extensive literature search. This revealed supporting evidence in the published literature for 10 out of the 12 entries (83.33%), thereby confirming the reliability of CADS in predicting novel drug combinations.

Notably, given that the DrugComb dataset used for training was updated in 2021 (Zheng *et al.* 2021), 40% confirmed novel drug combination entries were reported after this time point. This observation demonstrates the extensibility of the CADS, indicating its capability to discover novel drug combination-cell line causal relationships from a limited dataset.

3.4 Interpretability

In this section, we aimed to identify the key features of CADS that enable enhanced prediction accuracy.

3.4.1 Cell embedding

Based on the critical importance of cell line characterization data for determining whether drug combinations exhibit synergistic effects on specific cell lines, we are particularly interested in whether integrating CADS can enhance the model's capacity to capture cell line-specific features. In other words, we seek to investigate whether incorporating causal inference mechanisms enables the baseline models to generate more comprehensive representations of cell line information. To address this, we

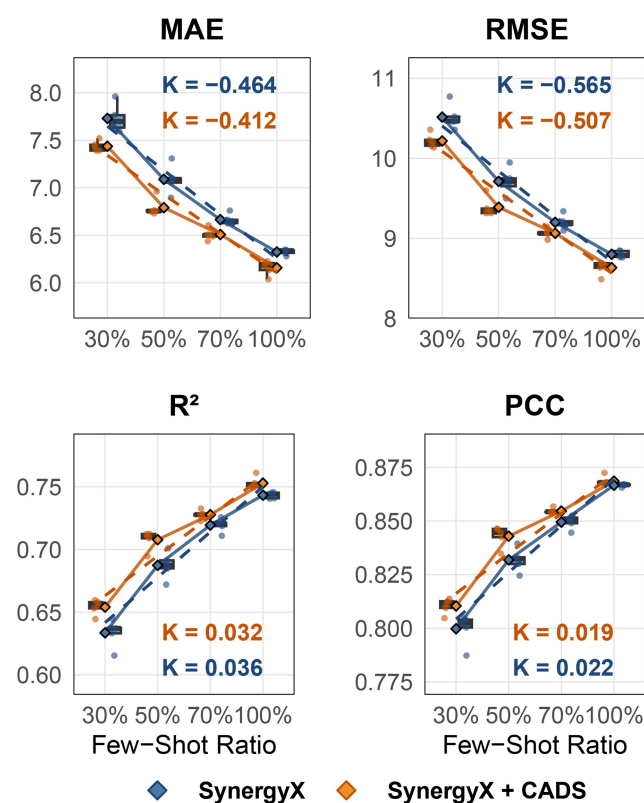


Figure 3 The Paired Comparison Plot illustrates four evaluation metrics from five sets of replicate experiments, comparing SynergyX+CADS and SynergyX under varying training set sizes for S-Score prediction.

extracted the final-layer representational vectors of 141 cell lines from the five-independent experiment cell line encoder in SynergyX—both with and without CADS integration—at the stage immediately prior to drug combination-cell line interaction. Fig. 4A–E presents the UMAP visualizations comparative analysis derived from five-independent replicate experiments. Based on consistent parameter configurations per batch, as shown in Fig. 4F, we observed increased Mean Within-Class Distance (Schroff *et al.* 2015) in every replicate experiment following CADS integration, indicating that CADS enhances feature separability in cell line representations.

3.4.2 General causal genes

To assess whether CADS isolates cell line-invariant causal information, we performed pan-cancer analysis of context-agnostic driver genes. Specifically, we computed the mean value and standard deviation of causal scores across 14 749 genes in a panel of 141 cell lines. Figure 5 displays the distribution of mean and variance for causal scores across genes. For comparative purposes, we adopted the housekeeping genes (Eisenberg and Levanon 2013) as the reference set. As illustrated, most housekeeping genes exhibit causal scores clustered between 0.45 and 0.55 with narrow variance, indicating that genes demonstrating low inter-cell-line variability, such as housekeeping genes, exert weaker effects on pharmacological synergy determination—a finding consistent with our initial hypothesis. We further investigated pan-causal genes in Appendix Section E.3.1 and E.3.2, available as [supplementary data](#) at Bioinformatics online.

Table 2 Novel drug combinations.

Cell line	Tissue	Drug A	Drug B	Predicted S-Score	Supporting Publications	Year
HCT-116	Colon	BI-2536	MK-2206	63.44± 2.00	PMID: 27699933 (Nonomiya et al. 2016)	2016
		AZD6738	AZD7762	62.29± 5.20	PMID: 36230796 (Deng et al. 2022)	2022
		MK-1775	Vincristine	60.51± 5.37	PMID: 39464635 (Song et al. 2024)	2024
		Erlotinib	Vincristine	59.64± 5.08	PMID: 40427071 (Li et al. 2025)	2025
A549	Lung	BI-2536	Dasatinib	52.56± 5.46	PMID: 32520674 (Zhang et al. 2021)	2021
		RAF265	Selumetinib	50.15± 6.01	PMID: 25199829 (Lamba et al. 2014)	2013
		AZD2014	AZD3463	49.95± 3.64	–	–
		BI-2536	RAF265	49.65± 6.27	(Weekes and Hidalgo 2010)	2010
MDA-MB-231	Breast	Bortezomib	Fludarabine	47.55 ± 2.83	–	–
		Bortezomib	Navitoclax	46.92± 4.14	PMID: 31370269 (Lim et al. 2019)	2019
		Bortezomib	Erlotinib	46.84± 2.75	PMID: 24747441 (Seguin et al. 2014)	2014
		Bortezomib	Doxorubicin	45.22± 4.87	PMID: 21147690 (Irvin et al. 2010)	2010

The table presents the mean values, standard deviations, and relevant references for S-Score predictions of novel drug combinations, derived from five sets of replicate experiments using SynergyX+CADS.

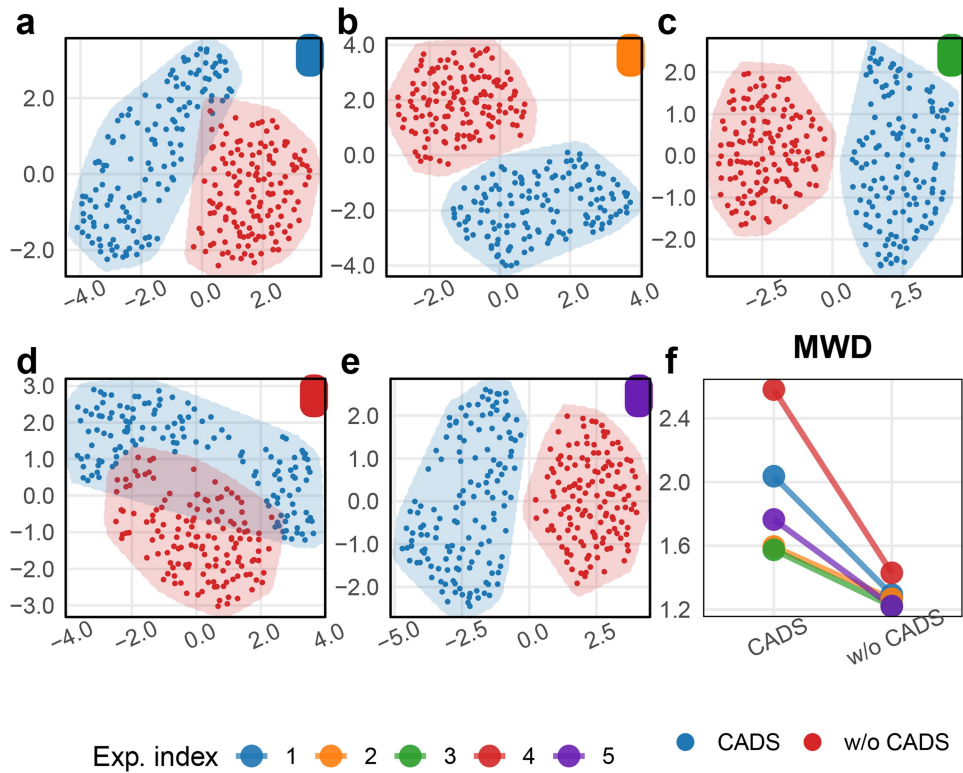


Figure 4 (a–e) The UMAP visualization illustrates the distribution of 141 cell line representation vectors from the causal component, derived from five sets of replicate experiments using SynergyX+CADS and SynergyX for S-Score prediction. (f) The Paired Comparison Plot illustrates the Mean Within-Class Distance (Schroff et al. 2015) across five groups of representation distributions.

3.4.3 Validation of specific causal genes

To validate the capability of CADS using external knowledge bases, we assessed the homogeneity and heterogeneity of causal genes influencing distinct drug combinations. We selected the two drug combinations with the broadest cellular coverage: (Ruxolitinib + Vismodegib) (89 of 141 cell lines) and (Crizotinib + Lenalidomide) (87 of 141 cell lines). Subsequently, gene-level attribution scores for both drug combinations across distinct cell lines were computed using Integrated Gradients

(Sundararajan et al. 2017). Mean contribution values were derived and subjected to Gene Set Enrichment Analysis (GSEA) (Korotkevich et al. 2016) in KEGG_LEGACY and KEGG_MEDICUS subsets (n = 369) in the MSigDB database. Figure 6 illustrates the top 10 enriched pathways based on the Normalized Enrichment Score (NES) for the two drug combinations.

Regarding the Crizotinib + Lenalidomide combination, Lenalidomide enhances immunoglobulin secretion in primary human B lymphocytes through activation of phosphatidylinositol

3-kinase δ (PI3K- δ)-dependent signaling cascades, a mechanism critical for its immunomodulatory effects in lymphoid malignancies (Lapalombella et al. 2010). Crizotinib exerts stronger antagonistic effects on receptor tyrosine kinase (RTK)-centric pathways, directly inhibiting ALK/MET-driven PI3K/RAS/ERK signal transduction to induce apoptosis and suppress proliferation in ALK-rearranged non-small cell lung cancer (NSCLC) cells (Shaw et al. 2013). Among the top 10 enriched pathways, RTK and PIK3 related pathways accounted for 30% (3/10), collectively demonstrating that CADS successfully captured core genetic determinants of this drug combination.

Regarding the Ruxolitinib + Vismodegib combination, several identified pathways demonstrate varying degrees of association with Ruxolitinib's mechanism of action. Notably, the IL-6 family to JAK-STAT signaling pathway is directly implicated, as Ruxolitinib potentially suppresses JAK-STAT activation downstream of IL-6 and related cytokines. The Hedgehog signaling pathway directly associated with Vismodegib exhibited a NES of 1.212 with an adjusted p-value > 0.05, indicating lack of statistical significance under conventional thresholds. This finding

indicates that CADS successfully delineated core genetic determinants influencing Ruxolitinib's mechanism, while partially underrepresenting Vismodegib-associated pathways, demonstrating a moderate feature selection bias in the model's analytical framework. We further analyzed this phenomenon in the Appendix Section E.3.3, available as supplementary data at Bioinformatics online.

This comparative analysis of causal genes between the two drug combinations validates that CADS effectively discriminates combinatorial regulatory mechanisms by revealing heterogeneous drivers underlying convergent phenotypes, thereby establishing its novel capability to decipher context-specific synergistic logic rather than merely memorizing generalizable gene sets. The consistency between the model-identified pathways and the experimentally established mechanisms of action documented in external databases serves as a validation of CADS's prediction reliability.

4 Conclusion

In this study, we introduced CADS, a novel framework for synergistic drug combination prediction. The model leverages causal inference principles to decouple biologically causal genes that critically influence drug response from trivial genetic associations within cell line profiles, thereby enhancing baseline predictive architectures. Experimental validation demonstrated concurrent improvements in both predictive accuracy and robustness against biological noise upon integration of the CADS module. Ablation studies confirmed CADS's capability to accurately isolate mechanistically relevant causal features, while representation analysis revealed that refined cell line embeddings underlie its performance gains. Biological interpretability analyses elucidated CADS's capacity to extract both pan-causal genes governing broad therapeutic responses and combination-specific mediators of synergistic effects.

Our analysis indicates that the "dominant-drug effect" is stochastic, with fluctuations in attention scores directly correlating to pathway sensitivity (e.g., in the Ruxolitinib-Vismodegib

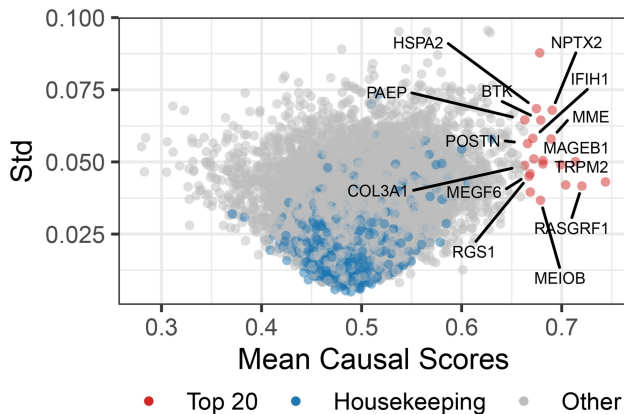


Figure 5 The volcano plot illustrates the mean and standard deviation of causal scores for 14,749 genes across 141 cell lines, derived from five sets of replicate experiments using SynergyX+CADS to predict S-Score.

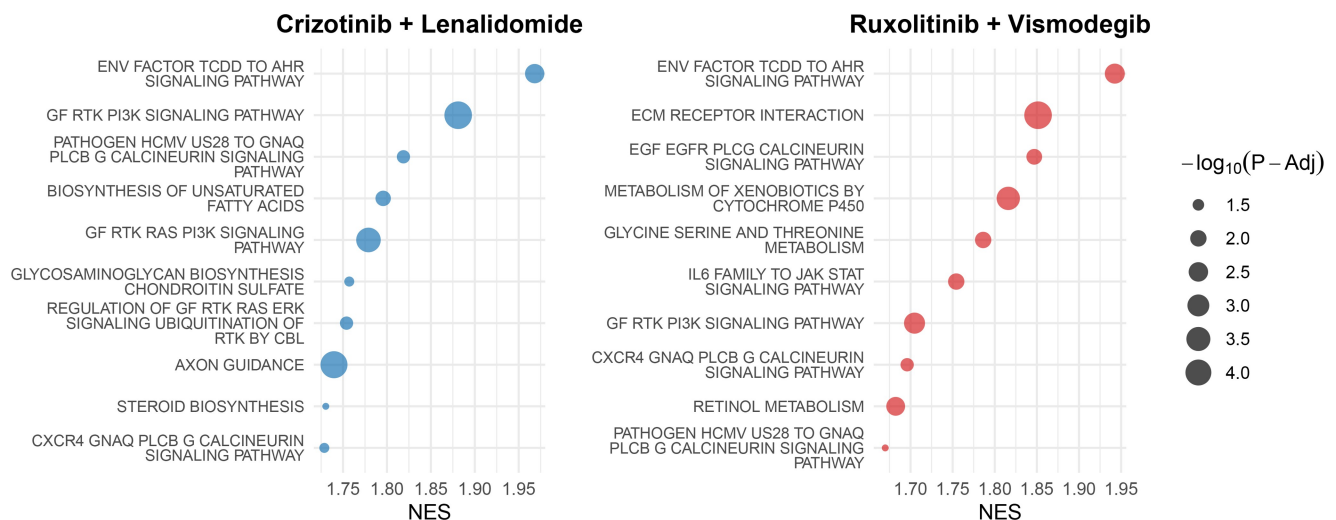


Figure 6 The bubble plot illustrates the top 10 enriched pathways identified through GSEA analysis of causal genes associated with two drug combinations, utilizing KEGG pathway sets ($n = 369$).

combination). Rather than enforcing a rigid symmetry, CADS supports a “contribution-aware” strategy: users can select specific experimental replicates based on drug contribution scores to optimize the identification of mechanisms of action.

Finally, we acknowledge potential threats to validity. Regarding internal validity, the high dimensionality and inherent noise in genomic data, coupled with the sparsity of available drug synergy labels, may pose challenges to the stability of causal discovery. Regarding external validity, our study relies on in vitro cell-line benchmarks; thus, generalizing these findings to in vivo clinical settings remains a challenge due to the complex tumor microenvironments that are not fully replicated in cell cultures.

Collectively, CADS establishes a paradigm for context-aware synergy prediction that transcends conventional feature correlation approaches, offering both computational efficacy and mechanistic interpretability for combinatorial drug discovery.

Author contributions

Huaiwu Zhang (Data curation [lead], Formal analysis [lead], Investigation [lead], Methodology [lead], Software [lead], Validation [lead], Visualization [lead], Writing—original draft [lead]), Xinliang Sun (Data curation [supporting], Formal analysis [supporting], Investigation [supporting], Methodology [supporting], Resources [supporting], Software [supporting], Validation [supporting], Visualization [supporting], Writing—review & editing [supporting]), Jianxin Wang (Conceptualization [equal], Investigation [supporting], Methodology [supporting], Project administration [supporting], Resources [supporting], Supervision [supporting], Writing—review & editing [supporting]), Min Li (Conceptualization [equal], Funding acquisition [equal], Resources [lead], Supervision [lead], Writing—review editing [supporting]), and Jing Tang (Conceptualization [lead], Funding acquisition [lead], Methodology [lead], Project administration [lead], Resources [lead], Supervision [lead], Writing—review & editing [lead])

Supplementary material

[Supplementary material](#) is available at *Bioinformatics* online.

Conflict of interests

None declared.

Funding

This work was supported by the Academy of Finland [No. 351165 and 357952].

Data availability

The data and source code of CADS are available at: <https://github.com/HuaiwuZhang/causalDC>.

References

- Ahmed SF, Alam MSB, Hassan M *et al*. Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artif Intell Rev* 2023;**56**:13521–617.
- Al-Lazikani B, Banerji U, Workman P. Combinatorial drug therapy for cancer in the post-genomic era. *Nat Biotechnol* 2012;**30**:679–92.
- Broach JT, Throner J. High-throughput screening for drug discovery. *Nature* 1996;**384**:14–6.
- Bulusu KC, Guha R, Mason DJ *et al*. Modelling of compound combination effects and applications to efficacy and toxicity: state-of-the-art, challenges and perspectives. *Drug Discov Today* 2016;**21**:225–38.
- Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and data mining*, p. 785–794. New York, NY, USA: ACM, 2016.
- Deng S, Vlatkovic T, Li M *et al*. Targeting the DNA damage response and DNA repair pathways to enhance radiosensitivity in colorectal cancer. *Cancers (Basel)* 2022;**14**:4874.
- Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet* 2013;**29**:569–74.
- Gilad Y, Gellerman G, Lonard DM *et al*. Drug combination in cancer treatment—from cocktails to conjugated combinations. *Cancers (Basel)* 2021;**13**:669.
- Gong L, Wei X, Ji H *et al*. A graph-based multi-dimensional interaction network for drug-drug interaction prediction. *IEEE J Biomed Health Inform* 2025;PP:1–14. <https://doi.org/10.1109/JBHI.2025.3575012>
- Guo Y, Hu H, Chen W *et al*. SynergyX: a multi-modality mutual attention network for interpretable drug synergy prediction. *Brief Bioinform* 2024;**25**:bbae015.
- He L, Tang J, Andersson EI *et al*. Patient-customized drug combination prediction and testing for T-cell prolymphocytic leukemia patients. *Cancer Res* 2018;**78**:2407–18.
- Hu J, Gao J, Fang X *et al*. DTSyn: a dual-transformer-based neural network to predict synergistic drug combinations. *Brief Bioinform* 2022;**23**:bbac302.
- Huang K, Xiao C, Glass L *et al*. Explainable substructure partition fingerprint for protein, drug, and more. In: *NeurIPS Learning Meaningful Representation of Life Workshop*, Vancouver, Canada: NeurIPS, 2019.
- Irvin WJ, Jr, Orlowski RZ, Chiu W-K *et al*. Phase II study of bortezomib and pegylated liposomal doxorubicin in the treatment of metastatic breast cancer. *Clin Breast Cancer* 2010;**10**:465–70.
- Jeon M, Kim S, Park S *et al*. In silico drug combination discovery for personalized cancer therapy. *BMC Syst Biol* 2018;**12**:16–67.
- Jiang P, Huang S, Fu Z *et al*. Deep graph embedding for prioritizing synergistic anticancer drug combinations. *Comput Struct Biotechnol J* 2020;**18**:427–38.
- Kim TK. T test as a parametric statistic. *Kor J Anesthesiol* 2015;**68**:540–6.
- Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: *International Conference on Learning Representations, Toulon, France*, 2017.
- Korotkevich G, Sukhov V, Budin N *et al*. Fast gene set enrichment analysis. biorxiv, 060012, 2016. Preprint: not peer-reviewed.

- Kuru HI, Tastan O, Cicek AE. MatchMaker: a deep learning framework for drug synergy prediction. *IEEE/ACM Trans Comput Biol Bioinform* 2022;**19**:2334–44.
- Lamba S, Russo M, Sun C *et al*. RAF suppression synergizes with MEK inhibition in KRAS mutant cancer cells. *Cell Rep* 2014;**8**:1475–83.
- Lapalombella R, Andritsos L, Liu Q *et al*. Lenalidomide treatment promotes CD154 expression on CLL cells and enhances production of antibodies by normal B cells through a PI3-kinase-dependent pathway. *Blood* 2010;**115**:2619–29.
- Li Y-C, Xiong Y-M, Long Z-P *et al*. ML210 antagonizes ABCB1-not ABCG2-mediated multidrug resistance in colorectal cancer. *Biomedicines* 2025;**13**:1245.
- Lim B, Greer Y, Lipkowitz S *et al*. Novel apoptosis-inducing agents for the treatment of cancer, a new arsenal in the toolbox. *Cancers (Basel)* 2019;**11**:1087.
- Lin W, Wu L, Zhang Y *et al*. An enhanced Cascade-based deep forest model for drug combination prediction. *Brief Bioinform* 2022;**23**:bbab562.
- Liu T, Chu T, Luo X *et al*. Building a unified model for drug synergy analysis powered by large language models. *Nat Commun* 2025;**16**:4537.
- Liu X, Song C, Liu S *et al*. Multi-way relation-enhanced hypergraph representation learning for anti-cancer drug synergy prediction. *Bioinformatics* 2022;**38**:4782–9.
- Noble WS. What is a support vector machine? *Nat Biotechnol* 2006;**24**:1565–7.
- Nonomiya Y, Noguchi K, Tanaka N *et al*. Effect of AKT 3 expression on MYC-and caspase-8-dependent apoptosis caused by polo-like kinase inhibitors in HCT 116 cells. *Cancer Sci* 2016;**107**:1877–87.
- Partin A, Brettin TS, Zhu Y *et al*. Deep learning methods for drug response prediction in cancer: predominant and emerging trends. *Front Med (Lausanne)* 2023;**10**:1086097.
- Plana D, Palmer AC, Sorger PK. Independent drug action in combination therapy: implications for precision oncology. *Cancer Discov* 2022;**12**:606–24.
- Preuer K, Lewis RPI, Hochreiter S *et al*. DeepSynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 2018;**34**:1538–46.
- Rafiei F, Zeraati H, Abbasi K *et al*. DeepTraSynergy: drug combinations using multimodal deep learning with transformers. *Bioinformatics* 2023;**39**:btad438.
- Sadybekov AV, Katritch V. Computational approaches streamlining drug discovery. *Nature* 2023;**616**:673–85.
- Schroff F, Kalenichenko D, Philbin J. FaceNet: a unified embedding for face recognition and clustering. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 815–823. Boston, MA, USA: IEEE, 2015.
- Seguin L, Kato S, Franovic A *et al*. An integrin $\beta 3$ -KRAS-RaB complex drives tumour stemness and resistance to EGFR inhibition. *Nat Cell Biol* 2014;**16**:457–68.
- Shaw AT, Kim D-W, Nakagawa K *et al*. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med* 2013;**368**:2385–94.
- Song G, Liu J, Tang X *et al*. Cell cycle checkpoint revolution: targeted therapies in the fight against malignant tumors. *Front Pharmacol* 2024;**15**:1459057.
- Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. In: *International Conference on Machine Learning*, p. 3319–3328. Sydney, Australia: PMLR, 2017.
- Veličković P, Cucurull G, Casanova A *et al*. Graph attention networks. In: *International Conference on Learning Representations (ICLR 2018)*, Vancouver, Canada.
- Wang J, Liu X, Shen S *et al*. DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Brief Bioinform* 2022;**23**:bbab390.
- Weekes CD, Hidalgo M. Targeted therapeutics in cancer treatment. In: *Principles of Anticancer Drug Development*, p. 403–461. NewYork: Springer, 2010.
- Wu L, Wen Y, Leng D *et al*. Machine learning methods, databases and tools for drug combination prediction. *Brief Bioinform* 2022;**23**:bbab355.
- Xu M, Zhao X, Wang J *et al*. DFFNDDS: prediction of synergistic drug combinations with dual feature fusion networks. *J Cheminform* 2023;**15**:33.
- Yang J, Xu Z, Wu WKK *et al*. GraphSynergy: a network-inspired deep learning model for anticancer drug combination prediction. *J Am Med Inform Assoc* 2021;**28**:2336–45.
- Yu J, Chen N, Cao L *et al*. Synergizing anti-cancer drug combinations with dual-view hypergraph representation fusion. *IEEE J Biomed Health Inform* 2026;**30**:3643–50. <https://doi.org/10.1109/JBHI.2024.3511657>
- Yu L, Xia M, An Q. A network embedding framework based on integrating multiplex network for drug combination prediction. *Brief Bioinform* 2022;**23**:bbab364.
- Zagidullin B, Aldahdooh J, Zheng S *et al*. DrugComb: an integrative cancer drug combination data portal. *Nucleic Acids Res* 2019;**47**:W43–51.
- Zhang H, Tang J. CADs: causal inference for dissecting essential genes to predict drug synergy. In: Tang J, Lai X, Cai Z, Peng W, Wei Y (eds.), *Bioinformatics Research and Applications, Volume 15756 of Lecture Notes in Computer Science*. Singapore: Springer, 2025, 288–300. https://doi.org/10.1007/978-981-95-0698-9_24
- Zhang L, Kang X, Yang X *et al*. MOSAIC: a multi-granularity cross-modal framework for predicting synergistic drug combinations in personalized healthcare. *IEEE J Biomed Health Inform* 2025;PP:1–12. <https://doi.org/10.1109/JBHI.2025.3605313>
- Zhang S, Chen Y, Tian C *et al*. Dual-target inhibitors based on BRD4: novel therapeutic approaches for cancer. *Curr Med Chem* 2021;**28**:1775–95.
- Zhao X, Xu J, Shui Y *et al*. PermutedDDS: a permutable feature fusion network for drug-drug synergy prediction. *J Cheminform* 2024;**16**:41.
- Zheng S, Aldahdooh J, Shadbahr T *et al*. DrugComb update: a more comprehensive drug sensitivity data repository and analysis portal. *Nucleic Acids Res* 2021;**49**:W174–W184.