

ParsVNN: parsimony visible neural networks for uncovering cancer-specific and drug-sensitive genes and pathways

Xiaoqing Huang¹, Kun Huang¹, Travis Johnson¹, Milan Radovich², Jie Zhang^{3,*}, Jianzhu Ma^{4,*} and Yijie Wang^{5,*}

¹Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, IN 46202, USA, ²Division of General Surgery, Department of Surgery, Indiana University School of Medicine, Indianapolis, IN 46202, USA, ³Department of Medical and Molecular Genetics, Indiana University, Indianapolis, IN 46202, USA, ⁴Institute for Artificial Intelligence, Peking University, China and ⁵Department of Computer Science, Indiana University, Bloomington, IN 47408, USA

Received March 31, 2021; Revised September 07, 2021; Editorial Decision October 04, 2021; Accepted October 08, 2021

ABSTRACT

Prediction of cancer-specific drug responses as well as identification of the corresponding drug-sensitive genes and pathways remains a major biological and clinical challenge. Deep learning models hold immense promise for better drug response predictions, but most of them cannot provide biological and clinical interpretability. Visible neural network (VNN) models have emerged to solve the problem by giving neurons biological meanings and directly casting biological networks into the models. However, the biological networks used in VNNs are often redundant and contain components that are irrelevant to the downstream predictions. Therefore, the VNNs using these redundant biological networks are overparameterized, which significantly limits VNNs' predictive and explanatory power. To overcome the problem, we treat the edges and nodes in biological networks used in VNNs as features and develop a sparse learning framework ParsVNN to learn parsimony VNNs with only edges and nodes that contribute the most to the prediction task. We applied ParsVNN to build cancer-specific VNN models to predict drug response for five different cancer types. We demonstrated that the parsimony VNNs built by ParsVNN are superior to other state-of-the-art methods in terms of prediction performance and identification of cancer driver genes. Furthermore, we found that the pathways selected by ParsVNN have great potential to predict clinical

outcomes as well as recommend synergistic drug combinations.

INTRODUCTION

For decades of cancer study, one of the most striking findings has been the extreme genetic heterogeneity among cancer patients (1–4). The heterogeneity of tumor cells poses a fundamental challenge for predicting the clinical response to therapeutic agents (5–9). To address this problem, there has been great interest recently to apply deep learning to model the complexity of cancer mutations on various cancer types (10–13). Despite the improved performance, deep learning in cancer still faces a critical challenge that they are still 'black box' models and are fundamentally more difficult to interpret than classical statistical models (14–16). Although they are accurate, they provide no meaningful insights about how their decisions are made. Such models, while undoubtedly useful, are insufficient in cancer studies for which clinicians need to understand the mechanisms underlying the predictions.

To address this problem, a new type of interpretable deep learning model emerged that coupled the neural network architecture with the hierarchical structure of a cell. Ma *et al.* (17) first proposed a visible neural network (VNN), named Dcell, to accurately predict the impact of genetic mutations on cellular growth response by directly mapping the neurons of a deep neural network into a large hierarchy of known and putative molecular components and pathways. It has been demonstrated that the VNN model could gain better interpretability than the conventional deep learning model by assigning biological meanings to the neurons and edges in the neural networks. This seminal research has spurred the development of other VNN models that cast

*To whom correspondence should be addressed. Tel: +1 812 856 9664; Email: yijwang@iu.edu
Correspondence may also be addressed to Jianzhu Ma. Email: majianzhu@pku.edu.cn
Correspondence may also be addressed to Jie Zhang. Tel: +1 317 274 2839; Email: jizhan@iu.edu

different types of biological networks into the architecture of deep learning models. For instance, Wang *et al.* (18) embedded the structure of the gene regulatory network derived from multi-omics data into a large multilevel deep learning model. The resulting model was adopted to impute intermediate phenotypes and provide forward predictions from genotypes to traits. DrugCell (19) was developed by combining a VNN model with a fully connected neural network model to simulate the response of cancer cells to therapeutic chemical compounds by guiding the neural network architecture with the hierarchical structure of a tumor cell. Very recently, another VNN model named P-NET (20) was designed to predict treatment resistance in prostate cancer patients using a biologically informed architecture that integrated mutation, copy number, methylation and gene expression information.

However, a major drawback of the VNN models is that the biological hierarchies used in VNN models might have redundant nodes and edges that are irrelevant to the downstream prediction tasks, which significantly limits the VNN models' explanatory power. These uninformative nodes and edges also contribute to the 'curse of dimensionality' and overfitting associated with highly nonlinear models like deep neural networks. Specifically, in order to incorporate rich prior information, the VNN model typically utilizes a general biological hierarchy that is agnostic to the downstream prediction tasks, which implies that some functional components in the biological hierarchy might not involve in the biological process related to the prediction task. However, the conventional learning algorithm used by the VNN model cannot distinguish these redundant functional components and therefore enforces them to make contributions. Such a phenomenon would lead the VNN model to generate misleading interpretations since even uninformative network components still contribute to the prediction task. Postprocessing procedures have been proposed to alleviate the problem by ranking functional components by their numerical contributions to the prediction accuracy (17,19,20). However, they cannot fundamentally resolve the problem because functional components that do not make biological contributions might make great numerical contributions. Furthermore, the redundant functional components result in overparameterized neural network architectures and thus require a lot of training data to avoid overfitting.

To address the above problems in VNNs, here we introduce ParsVNN, a method to prune redundant components in the VNN model to make its biological hierarchy simple and specific to the prediction task. ParsVNN starts from a VNN model with a general biological hierarchy and uses the training data of a specific prediction task to guide the pruning procedure. As a result, redundant components in the general biological hierarchy used in the VNN model that are less important to the prediction task will be pruned and the VNN model will become a parsimonious model with greater explanatory predictive power. ParsVNN addresses important challenges in building biologically interpretable deep learning models. It can be used to elucidate the interpretability of deep learning models that utilize biological networks in their architectures.

We applied ParsVNN to build cancer-specific VNN models to help us better understand the biological mechanisms

for five different cancers: stomach, breast, pancreatic, kidney and liver cancers. We use ParsVNN to prune the DrugCell model whose architecture mirrors the Gene Ontology (GO) hierarchy (21,22) of a human cell using the cancer-specific training samples collected from the Cancer Therapeutics Response Portal (CTRP) v2 and the Genomics of Drug Sensitivity in Cancer (GDSC) database (23,24). To validate the predictive power of models generated by ParsVNN, we evaluate their prediction performance on the test samples. We compared ParsVNN with several state-of-the-art methods and found that the models generated by ParsVNN have extremely simple architectures compared to the original DrugCell (19), and surprisingly, most of them achieve significantly better test accuracy. In addition, we examined the explanatory power of the models generated by ParsVNN. We found that the genes remaining in these ParsVNN models consist of more cancer driver genes (25), co-occurrence genes and mutually exclusive genes than the competing methods. Furthermore, we discovered that the functional modules in biological hierarchies recognized by ParsVNN can help predict clinical outcomes for different cancers. Finally, we demonstrated that we can analyze the subsystems identified by ParsVNN to suggest synergistic drug combinations.

MATERIALS AND METHODS

ParsVNN overview

The main idea of ParsVNN is to use cancer-specific drug response data to prune the redundant components in the biological hierarchy used in a VNN model. The resulting biological network in the pruned model will be simple and important for explaining drug response in cancers. To accomplish this, ParsVNN requires three components: (i) a VNN model with the general biological hierarchy to begin with (Figure 1A); (ii) the cancer-specific drug response samples to guide the pruning process (Figure 1B); and (iii) an algorithmic technique guaranteeing convergence of the pruning procedure. Below we provide a basic explanation of these three components. The details of the method and its mathematical underpinning are described in this section.

As shown in Figure 1A, the VNN model uses a biological hierarchy as its architecture. Because the biological hierarchy is agnostic to the cancer type we aim to model, some of the nodes and edges in it might be redundant and irrelevant to the underlying mechanism. The drug response training samples for a specific cancer (Figure 1B) can help to guide ParsVNN to prune those redundant components. After applying ParsVNN, a parsimony VNN model will be obtained (Figure 1C.1). Then, we can further analyze the parsimony architecture to interpret the predictive power of the obtained model in Figure 1C.1. Specifically, we can identify the essential genes for the drug response (Figure 1C.2) and important pathways that lead to different patient survival rates (Figure 1C.3). Furthermore, by analyzing the parsimony architecture in Figure 1C.4, new drug combinations with synergistic effective would be suggested.

Overall, as shown in Figure 1, starting from a VNN model ParsVNN treats each edge weight as a feature of the VNN model and performs sparse learning to simultaneously enhance prediction accuracy as well as select

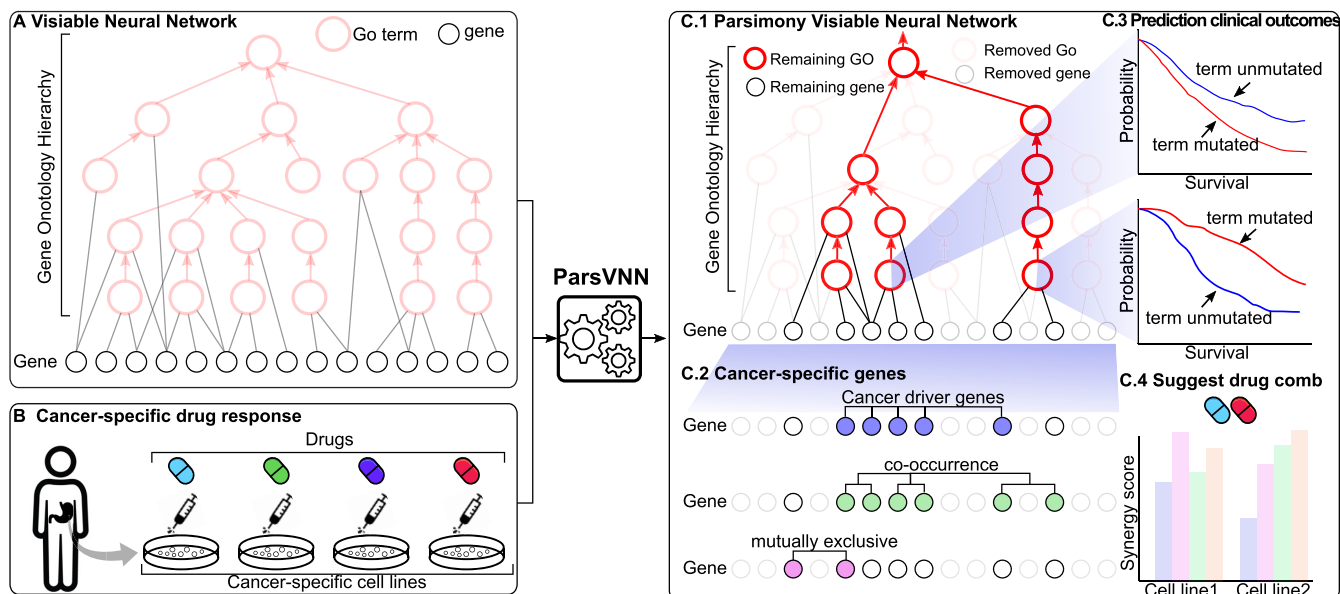


Figure 1. The overview of ParsVNN. (A) A VNN model using a specific biological network as its architecture. (B) Cancer-specific drug response data to guide the pruning in ParsVNN. (C.1) The resulting parsimony VNN model after applying ParsVNN. (C.2) Genes identified by ParsVNN are enriched in cancer driver genes, co-occurrence genes and mutually exclusive genes. (C.3) Mutation of the GO terms remaining in panel (C.1) leads to different patient survival. (C.4) New drug combinations could be suggested by panel (C.1).

important features/edges, the latter removing redundant features/edges and allowing better interpretability for the downstream analysis (Figure 1C.1–C.4). Computationally, ParsVNN is formulated as a sparse learning problem, where ℓ_0 norm regularization is used to prune edges between genes and subsystems, while group lasso regularization is applied to remove edges between subsystems. The novel combination of ℓ_0 norm regularization and group lasso regularization makes the corresponding training objective nonconvex and NP-hard. We addressed this computational challenge by using a new cutting-edge technique known as proximal alternative linearized minimization (PALM) (26) as described in this section. The source code of ParsVNN is available at <https://github.com/EJIUB/ParsVNN>.

Visible neural network model

A visible neural network (VNN) model is a transparent deep learning model that aims to characterize the relationship between gene-level measurements and the corresponding phenotypic response in a cell. The architecture of the VNN model mirrors the hierarchical organization of molecular subsystems in a cell. Specifically, as shown in Figure 2, the VNN model consists of artificial neurons that represent genes (black color in Figure 2) and artificial neurons that represent molecular subsystems (pink color in Figure 2). As illustrated in Figure 2, we call the neurons that represent genes the ‘gene-neurons’ and the neurons that represent molecular subsystems the ‘subsystem-neurons’, respectively. We further use $\mathcal{V}^I = \{v_1^I, \dots, v_p^I\}$ to annotate all ‘gene-neurons’ and use $\mathcal{V}^S = \{v_1^S, \dots, v_q^S\}$ to annotate all ‘subsystem-neurons’. Notably, as illustrated in Figure 2, each ‘gene-neuron’ is a single artificial neuron, but each ‘subsystem-neuron’ is a group of neurons. The use of mul-

iple neurons to represent a subsystem acknowledges that molecular subsystems are often multifunctional, with states that are too complex to be described by a single neuron (19).

The wiring between ‘gene-neurons’ \mathcal{V}^I and ‘subsystem-neurons’ \mathcal{V}^S and connectivity between ‘subsystem-neurons’ \mathcal{V}^S are provided by biological prior knowledge. Specifically, if we know the gene annotation for each molecular subsystem, we can connect ‘gene-neurons’ to the corresponding ‘subsystem-neurons’. The black edges in Figure 2 represent such connectivity. Similarly, if we know the hierarchical organization of molecular subsystems, we are able to link ‘subsystem-neurons’. The pink edges in Figure 2 represent interactions between ‘subsystem-neurons’. Let E^I be the edge set including all edges between ‘gene-neurons’ \mathcal{V}^I and ‘subsystem-neurons’ \mathcal{V}^S and E^S be the edge set containing all interactions between ‘subsystem-neurons’ \mathcal{V}^S . We further define W_e^I , $e \in E^I$, as the weight of the edge e in E^I and define W_t^S , $t \in E^S$, as the weight matrix associated with the edge t in E^S .

Given n training samples $\{(x^i, y^i); i = 1, \dots, n\}$, where $x^i \in \mathbb{R}^p$ is the i th input vector for the VNN model and its element X_j^i , $j \in \{1, \dots, p\}$, is the measurement from gene j , and y^i is the corresponding phenotypic output, the VNN model can be trained by solving the following empirical risk minimization:

$$\min_{W^I, W^S} : \mathcal{L}(W^I, W^S) := \frac{1}{n} \sum_i L(y^i, f(W^I, W^S; x^i)), \quad (1)$$

where $W^I = \{W_e^I, e \in E^I\}$ and $W^S = \{W_t^S, t \in E^S\}$. L is the loss function and $f(W^I, W^S; x^i)$ yields the predicted phenotypic response \hat{y}^i for input x^i . We can use the off-the-shelf conventional optimization methods to minimize the loss function and learn the model parameters in W^I and W^S .

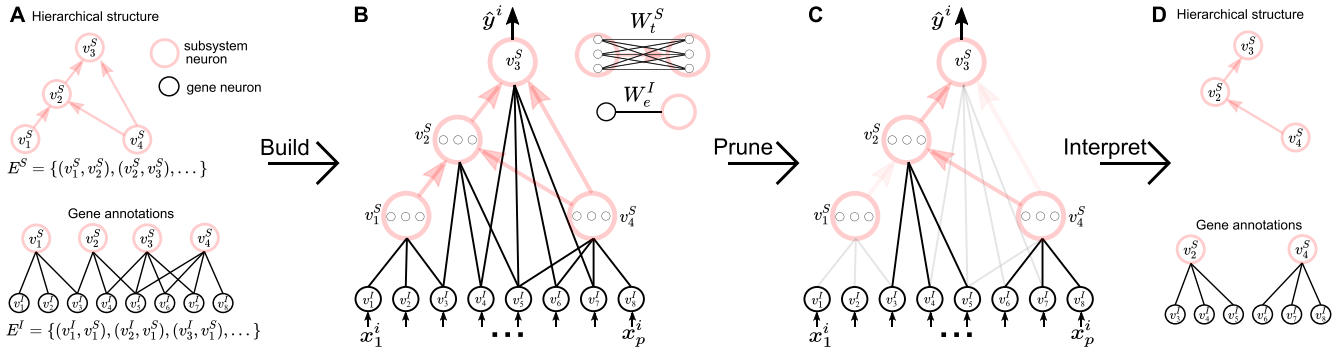


Figure 2. The overview of ParsVNN. (A) Biological knowledge is encoded in networks. (B) A VNN model embeds the networks in its architecture. (C) ParsVNN prunes the VNN model by removing less important edges for a specific prediction task. The resulting network is essential for the prediction task. (D) We can further analyze the components in the resulting network to study how they related to the prediction task.

Learning a parsimony VNN model

We hypothesize that ‘gene-neurons’ and ‘subsystem-neurons’ do not contribute equally to the predictive power of the VNN model. There must be a small portion of ‘gene-neurons’ and ‘subsystem-neurons’ that are crucial to the VNN model. Therefore, we propose to learn a parsimony VNN model that only retains the key ‘gene-neurons’ and ‘subsystem-neurons’ as shown in Figure 2. Furthermore, because the ‘input neurons’ and ‘subsystem neurons’ all have specific biological meanings, the remaining ‘gene-neurons’ and ‘subsystem-neurons’ can help us to better interpret the VNN model.

We propose to learn the VNN model with parsimonious structures by pruning less important weights in W^I and W^S in the VNN model. The pruning of the VNN model can be achieved by solving the following optimization problem:

$$\min_{W^I, W^S} : \mathcal{L}(W^I, W^S) + \Omega_\lambda(W^I) + \Gamma_\eta(W^S). \quad (2)$$

Comparing to Equation (1), we add two sparse-inducing penalty terms $\Omega_\lambda(W^I)$ and $\Gamma_\eta(W^S)$ to remove less important edges in E^I and E^S , respectively. Specifically, $\Omega_\lambda(W^I)$ is defined as

$$\Omega_\lambda(W^I) = \lambda \sum_{e \in E^I} \|W_e^I\|_0, \quad (3)$$

where λ is the regularization parameter. For each edge $e \in E^I$, $\|W_e^I\|_0$ is 1 when $W_e^I \neq 0$ and 0 otherwise. Minimizing Equation (3) tends to zero out the weight of the edge $e \in E^I$, which results in removing the edge between the corresponding ‘gene-neuron’ and ‘subsystem-neuron’. In other words, the penalty term in Equation (3) could help refine the gene annotation of each molecular subsystem as illustrated in Figure 2. We note that using ℓ_1 norm is also able to prune less important edges in the VNN model. We choose to use ℓ_0 norm because ℓ_0 norm has been proved to have better empirical performance on selecting sparse features (27,28). The other sparse-inducing penalty term $\Gamma_\eta(W^S)$ is defined as follows:

$$\Gamma_\eta(W^S) = \eta \sum_{t \in E^S} \|W_t^S\|_g, \quad (4)$$

where η is the regularization parameter that controls the strength of the penalty term. $\|W_t^S\|_g = \sqrt{\sum_{ij} W_t^S(i, j)^2}$ [$W_t^S(i, j)$ is the element in W_t^S] is the group lasso term that is applied to the groups of weights associated with the connection $t \in E^S$ between ‘subsystem-neurons’. Minimizing Equation (4) tends to zero out all weights in W_t^S resulting in the removal of the connection $t \in E^S$. Such removal can be interpreted as the interaction between the corresponding molecular subsystems is less important for the learning task in Equation (1).

Algorithm 1 The ParsVNN Algorithm

- 1: Initialize $(W^I)^0, \forall e, (W^S)^0, \forall t, \mathcal{L}_\eta^0 > 0$, and $\mathcal{L}_\lambda^0 > 0$.
- 2: for $k = 0, 1, 2, \dots$ do
- 3: $H^k = \nabla_{W^I} \mathcal{L}((W^I)^k)$.
- 4: Find $\mathcal{L}_\lambda = \min\{2^k \mathcal{L}_\lambda^{k-1}\}$.
- 5: s.t. $\|H^{k+1} - H^k\|_2 \leq \mathcal{L}_\lambda \| (W^I)^{k+1} - (W^I)^k \|_2, b \in \mathbb{Z}$.
- 6: $(W^I)^{k+1} \in T_{\frac{\mathcal{L}_\lambda}{\sqrt{2^k}}}(H^k), \forall e \in E^I$.
- 7: $F^k = \nabla_{W^S} \mathcal{L}((W^I)^{k+1}, (W^S)^k)$.
- 8: Find $\mathcal{L}_S = \min\{2^k \mathcal{L}_S^{k-1}\}$.
- 9: s.t. $\|F^{k+1} - F^k\|_2 \leq \mathcal{L}_S \| (W^S)^{k+1} - (W^S)^k \|_2, b \in \mathbb{Z}$.
- 10: $(W^S)^k \in \max(\|V_t^S\| - \frac{\mathcal{L}_S}{\eta}, 0) \frac{V_t^S}{\|V_t^S\|}, t \in E^S$.
- 11: $\mathcal{L}_\eta^k = \mathcal{L}_\eta$ and $\mathcal{L}_\lambda^k = \mathcal{L}_\lambda$.
- 12: end for

The ParsVNN algorithm

The optimization problem in Equation (2) is challenging to solve. Both the sparse-inducing penalty terms are not differentiable; therefore, conventional optimization techniques for regular deep learning models cannot be applied. We employ proximal alternative linearized minimization (PALM) algorithm (29) to solve the proposed formulation in Equation (2). PALM is designed to solve a general optimization problem formulated as

$$\min_{W^I, W^S} : F(W^I, W^S) + \Phi_I(W^I) + \Phi_S(W^S), \quad (5)$$

where $F(W^I, W^S)$ is a smooth function and $\Phi_I(W^I)$ and $\Phi_S(W^S)$ do not need to be convex or smooth but are only required to be lower semi-continuous. The PALM algorithm applies the proximal forward-backward algorithm (29) to optimize both W^I and W^S in an alternative manner. Specifically, at iteration k , the proximal forward-backward mappings of $\Phi_I(W^I)$ and $\Phi_S(W^S)$ for given $(W^I)^k$ and $(W^S)^k$ are

the solutions of the following subproblems, respectively:

$$(W^A)^{k+1} \in \min_{W^A} : \left\{ \frac{c^k}{2} \|W^A - U^k\|_F^2 + \Phi_I(W^A) \right\}, \quad (6a)$$

$$(W^S)^{k+1} \in \min_{W^S} : \left\{ \frac{d^k}{2} \|W^S - V^k\|_F^2 + \Phi_S(W^S) \right\}, \quad (6b)$$

where $U^k = (W^A)^k - (1/c^k)\nabla_{W^A} F((W^A)^k, (W^S)^k)$ and $V^k = (W^S)^k - (1/d^k)\nabla_{W^S} F((W^A)^{k+1}, (W^S)^k)$. c^k and d^k are positive real numbers and $\nabla_{W^A} F((W^A)^k, (W^S)^k)$ is the derivative of $F(W^A, (W^S)^k)$ with respect to W^A at point $(W^A)^k$ for fixed $(W^S)^k$ and $\nabla_{W^S} F((W^A)^{k+1}, (W^S)^k)$ is the derivative of $F((W^A)^{k+1}, W^S)$ with respect to W^S at point $(W^S)^k$ for fixed $(W^A)^{k+1}$. It has been proven that the sequence $\{(W^A)^k, (W^S)^k\}_{k \in \mathbb{N}}$ generated by PALM converges to a critical point (29).

Casting our proposed optimization problem (2) into the PALM framework (5) introduced above, we have $F(W^A, W^S) := \mathcal{L}(W^A, W^S)$, $\Phi_I(W^A) := \Omega_\lambda(W^A)$ and $\Phi_S(W^S) := \Gamma_\eta(W^S)$. It is easy to verify that $F(W^A, W^S)$, $\Phi_I(W^A)$ and $\Phi_S(W^S)$ satisfy the requirements of the PALM algorithm. Hence, we can apply the PALM algorithm to our problem as long as we can efficiently solve the proximal forward-backward mappings for our specific $\Phi_I(W^A)$ and $\Phi_S(W^S)$.

Putting $\Phi_I(W^A) := \Omega_\lambda(W^A)$ and $\Phi_S(W^S) := \Gamma_\eta(W^S)$ into Equations (6a) and (6b), we have the specific subproblem for our formation (2):

$$(W^A)^{k+1} \in \min_{W^A} : \left\{ \frac{c^k}{2} \|W^A - U^k\|_F^2 + \lambda \sum_{e \in E^I} \|W_e^A\|_0 \right\}, \quad (7a)$$

$$(W^S)^{k+1} \in \min_{W^S} : \left\{ \frac{d^k}{2} \|W^S - V^k\|_F^2 + \eta \sum_{t \in E^S} \|W_t^S\|_g \right\}. \quad (7b)$$

Both subproblems (7a) and (7b) have closed-form solutions as follows:

$$(W_e^A)^{k+1} \in T_{\sqrt{\frac{2\lambda}{c^k}}}(U_e^k), \quad \forall e \in E^I, \quad (8a)$$

$$(W_t^S)^{k+1} \in \max \left(\|V_e^k\| - \frac{2\eta}{d^k}, 0 \right) \frac{V_e^k}{\|V_e^k\|}, \quad t \in E^S. \quad (8b)$$

$T_c(\cdot)$ is the hard-thresholding operator (29). We skip the derivations of (8a) and (8b) because similar derivations can be found in (30).

We now have all the ingredients for our ParsVNN algorithm. Hence, we describe the VNN.Prune algorithm in Algorithm 1. The operations from line 3 to line 6 compute the solution for subproblem (7a), where line 4 and line 5 estimate the Lipschitz constant of $\nabla_{W^A} \mathcal{L}(W^A, \cdot)$. Similarly, the operations from line 7 to line 10 compute the solution for subproblem (7b), where line 8 and line 9 estimate the Lipschitz constant of $\nabla_{W^S} \mathcal{L}(\cdot, W^S)$.

Build cancer-specific ParsVNN models

We applied the ParsVNN algorithm to prune the DrugCell model to predict drug responses for specific cancer types.

Table 1. Details of the cancer-specific data

Cancer type	No. of samples	No. of cell types	No. of drugs
STAD	17 964	42	684
BRCA	24 928	57	684
PAAD	22 101	42	684
KICH/KIRC/KIRP	13 463	38	684
LIHC	12 550	26	684

Data. In order to build cancer-specific ParsVNN models, we collected cancer-specific data from the CTRP v2 and the GDSC database (23,24). We have collected data for five specific cancer types, which are stomach adenocarcinoma (STAD), breast invasive carcinoma (BRCA), pancreatic adenocarcinoma (PAAD), kidney chromophobe/kidney renal clear cell carcinoma/kidney renal papillary cell carcinoma (KICH/KIRC/KIRP) and liver hepatocellular carcinoma (LIHC). Table 1 provides an overview of each cancer dataset. For example, STAD-specific data include 17 964 cell line–drug pairs covering 42 stomach cancer cell lines and 684 drugs.

The DrugCell model. DrugCell is a VNN model that is built on the GO hierarchy and has been successfully applied to predict drug response and synergy (19). The architecture of DrugCell is shown in Figure 3A and has two main branches. The first branch is a conventional artificial neural network (ANN) that takes the Morgan fingerprint of a drug, a canonical vector representation of chemical structure (31), as the input. This branch intends to embed the chemical structure of drugs. The second branch is a VNN model, whose architecture is provided by the GO hierarchy of a cell. The GO hierarchy consists of 3008 GO terms (subsystems in Figure 1) and 2086 genes. There are 3176 interactions between GO terms, which represent the functional organization of the GO terms. There are 19 744 connections between GO terms and specific genes, which represent the gene annotation of each GO term (21,22). The VNN branch takes cancer cell-line genotype, which is represented by a binary vector registering the mutational status (1 = mutated, 0 = nonmutated) of each gene as input. The VNN branch aims to embed the genotype of different cancer cell lines. Overall, DrugCell integrates the embeddings of drug structures and genotypes and associates them with corresponding drug responses.

The ParsVNN model. The original architecture of a ParsVNN model is exactly the same as the DrugCell model as shown in Figure 3A, which has an ANN branch for drug embedding and a VNN branch for genotype embedding. We then apply the ParsVNN algorithm to train and prune the original ParsVNN model using cancer-specific training data. We train and prune edges in the VNN branch. For the ANN branch, we train the edge weights without pruning. After training, we obtained the cancer-specific ParsVNN model, where the VNN branch would have a cancer-specific architecture as shown in Figure 3A.

The competing models. We compared our cancer-specific ParsVNN models with models built by other methods. We

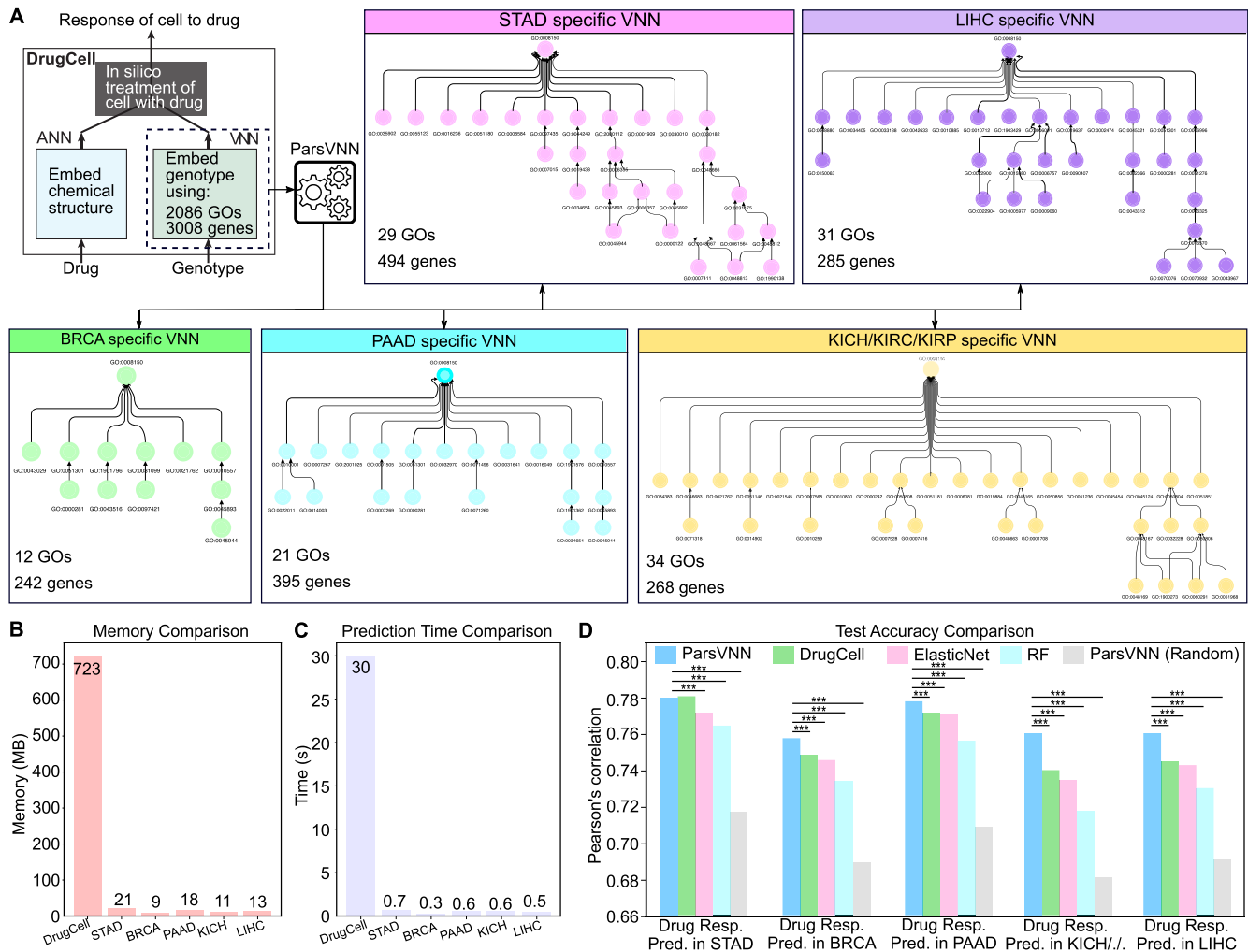


Figure 3. The architectures of five cancer-specific ParsVNN models and their performance on test samples. (A) The architecture of DrugCell. The resulting architectures of ParsVNN models for STAD, BRCA, PAAD, KICH/KIRC/KIPR and LIHC cancer types after applying ParsVNN to the original DrugCell model using cancer-specific training samples. (B) Memory comparison of DrugCell and the pruned models. (C) Prediction time comparison of DrugCell and the pruned models. Prediction time is the average forward passing time of the deep learning models for 10 000 inputs over 10 times. (D) Performance comparison between ParsVNN and other competing methods on drug response prediction for five different cancer types. RF stands for random forest. We use the Pearson's correlation between the predicted drug responses and the observed drug responses as the evaluation criterion. The asterisks (***) indicate the P -value < 0.0001 .

first compared with the original DrugCell models that are trained on cancer-specific samples. We also compared with ParsVNN (random) models, where the biological meanings of each neuron are randomly assigned. We compared with the models generated from the elastic net (ElasticNet), a state-of-the-art sparse learning technique that could also yield parsimony models and has been used in many previous approaches to drug response prediction (19,32–34). Last but not least, we compared with the random forest (RF) model. For the ElasticNet and RF models, we build the input feature set by stacking the Morgan fingerprint of drugs and the genotypes of cells. All competing models take the same input data. The difference among them is the way they extract features.

Training and evaluation. We use 80% of the cancer-specific samples for training and 20% samples for testing for each competing model. The hyperparameters for each competing

model are selected by 5-fold cross-validation on the training samples. We compared the competing models in terms of the prediction accuracy, which is estimated by Pearson's correlation between predicted drug response and the observed area under the curve values provided in (31).

The final models. The final cancer-specific models are built by using the entire cancer-specific samples and the selected hyperparameters.

RESULTS

Cancer-specific ParsVNN models have parsimonious architectures and improved performance

Following the 'Build cancer-specific ParsVNN models' section, we built five cancer-specific ParsVNN models and benchmarked them with other cancer-specific competing

models. First, we observed that the learned cancer-specific ParsVNN models are very compact as shown in Figure 3A. For example, the STAD-specific ParsVNN model has only 29 GO terms and 494 genes left, which indicates around 98% GO terms and 83% genes have been removed from the original DrugCell model. Similarly, all the remaining cancer-specific ParsVNN models have very few GO terms and genes left. Because the cancer-specific ParsVNN models are extremely compact, as shown in Figure 3B and C, the memory consumption of each model and the forward passing prediction time have been dramatically reduced.

We further found out that the obtained cancer-specific ParsVNN models although compact have better prediction performance. Specifically, we compared the performance of the cancer-specific ParsVNN models with other models (19,35) in terms of test accuracy. The test accuracy is measured by Pearson's correlation between the predicted drug responses and the ground truth drug responses. As illustrated in Figure 3D, the cancer-specific ParsVNN models achieve higher Pearson's correlation on all five cancer-specific datasets, except that DrugCell is slightly better than the ParsVNN model on the STAD dataset. We further tested whether the correlation achieved by ParsVNN is statistically better than others (36). It turns out that the correlation achieved by ParsVNN is significantly larger than the competing algorithms with P -values <0.0001 for BRCA, PAAD, KICH/KIRC/KIRP and LIHC cancer types (36). DrugCell achieves a slightly larger correlation than ParsVNN on STAD cancer type, but the difference is not significantly larger with a P -value of 0.68 (36), which indicates ParsVNN and DrugCell have similar test accuracy for STAD cancer data. The performance comparison indicates that the ParsVNN is able to prune redundant architectures of VNN models into compact ones and thus reduces the generalization error of the models.

Remarkably, the architectures in these five cancer-specific ParsVNN models (Figure 3A) are quite different, which confirms that drug responses are cancer type specific. In the following sections, we scrutinize these parsimonious architectures to see whether they can provide the biological insights to explain cancer-specific drug responses.

ParsVNN identifies cancer driver genes

To demonstrate ParsVNN models have better explanatory power, we studied the genes that remained in the parsimonious architectures of the ParsVNN models. It has been shown that cancer driver genes are the key factor to influence cancer-specific drug response (37). Therefore, we measure the overlap between the genes in the ParsVNN models and the cancer-specific driver genes reported by IntOGen pipeline (25). As illustrate in Figure 4A, the Venn diagrams show the overlap between genes identified by ParsVNN and ElasticNet and the cancer driver genes reported by IntOGen pipeline (25) for five different cancer types. Clearly, ParsVNN selected more cancer driver genes than ElasticNet. We further computed the odds ratios, 95% confidence intervals and P -values (H_0 : the odds ratio is ≤ 1) for the gene lists identified by both methods. As shown in Figure 4B, the gene lists identified by ParsVNN for STAD, BRCA and PAAD attain higher odds ratios than ElasticNet and the

gene lists identified by ElasticNet achieve higher odds ratios than ParsVNN. However, the gene lists identified by ParsVNN achieve narrower confidence intervals across all five cancer types, which suggests that we have more statistically precise odds ratios obtained by ParsVNN. Furthermore, the gene lists identified by ParsVNN also achieve lower P -values across all five cancer types, which indicates that ParsVNN provides stronger evidence to reject the null hypothesis (the odds ratio is ≤ 1). Furthermore, we studied the number of co-occurrence and mutually exclusive gene pairs within the selected gene set found by ParsVNN and ElasticNet. Co-occurrence and mutually exclusive are important mutational patterns that can help characterize specific cancers. For example, we use cBioPortal (38,39) to identify 75 224 and 360 co-occurrence and mutually exclusive gene pairs within the 494 genes found by ParsVNN for STAD cancer. However, there are only 578 and 2 co-occurrence and mutually exclusive gene pairs within the 45 genes found by ElasticNet. We have the same observation for the rest of the cancer types as shown in Figure 4C and D. Clearly, the genes selected by ParsVNN have richer mutational information than genes found by ElasticNet. The detailed numbers can be found in Supplementary Data.

These results indicate that ParsVNN has better explanatory power than ElasticNet, which is a state-of-the-art sparse learning technique used in many previous approaches to drug response prediction (19,32–34).

Cancer-specific subsystems identified by ParsVNN predict clinical outcomes

To illustrate that the subsystems identified by ParsVNN also have explanatory power, we explored the relationships between subsystems in ParsVNN models and clinical outcomes of cancer patients. We screened the leaf GO terms of each cancer-specific GO hierarchy and study the influence of the mutation of each leaf GO term with respect to cancer survival.

We first examined 15 leaf GO terms for GO hierarchy in the STAD-specific ParsVNN model. The cBioPortal's survival analysis tool and all STAD samples (40–45) are used to analyze each leaf GO term. Specifically, for each GO term, we split all STAD samples into two groups, where one group contains the samples that do not have any genes mutated in the GO term and the other group includes all samples that have at least one gene mutated in the GO term. We found that the mutations of 4 GO terms out of 15 lead to significantly different survival rates (P -value <0.01). The details of these four GO terms are listed in Figure 5F and the genes annotated in those GO terms can be found in Supplementary Figure S1. For example, we scrutinize GO:0016235 shown in Figure 5A. We found that most of the genes in the term are co-occurrence genes. Surprisingly, as illustrated in Figure 5A, patients with the mutation of the GO term have better survival rates compared to those without the mutation, which indicates that the mutation of the GO term is beneficial.

We investigated the six leaf GO terms in the GO hierarchy identified by BRCA-specific ParsVNN model. For each GO term, we did survival analysis on samples with the term mutated and samples with the term unmutated using

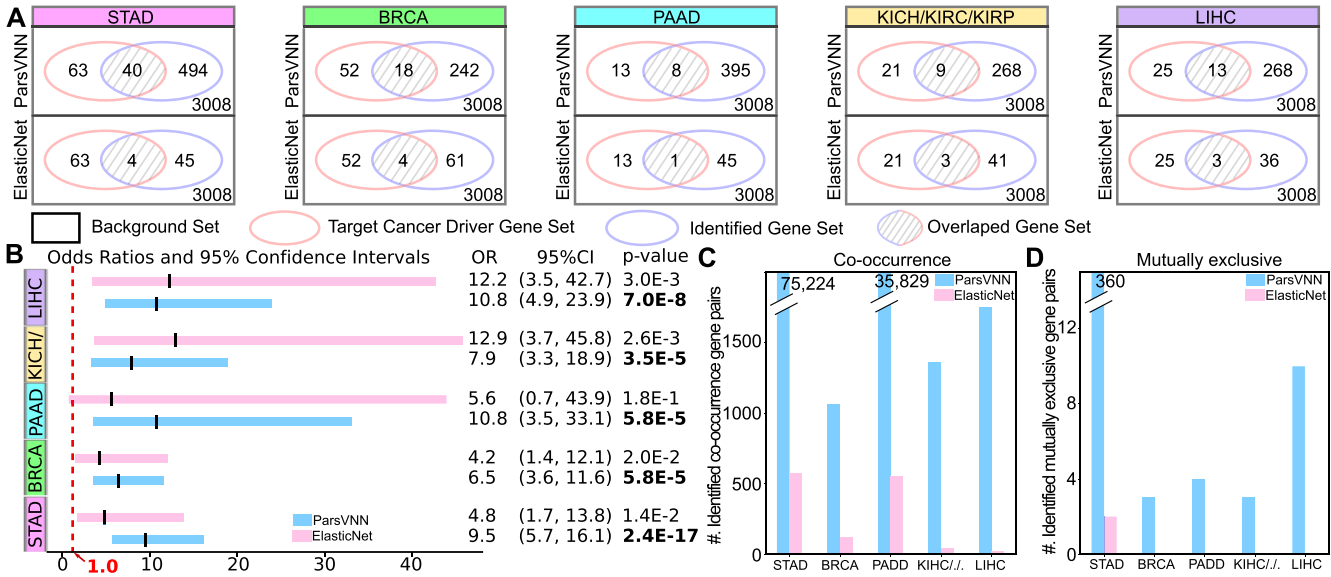


Figure 4. Comparison between genes selected by ParsVNN and ElasticNet. (A) Comparison of agreement with cancer driver genes using Venn diagrams. (B) Comparison of odds ratios, 95% confidence intervals and P-values (H0: the odds ratio is ≤ 1). OR stands for odds ratio and CI stands for confidence interval. (C) Comparison of the number of co-occurrence gene pairs in the selected gene sets of both methods. (D) Comparison of the number of mutually exclusive gene pairs in the selected gene sets of both methods.

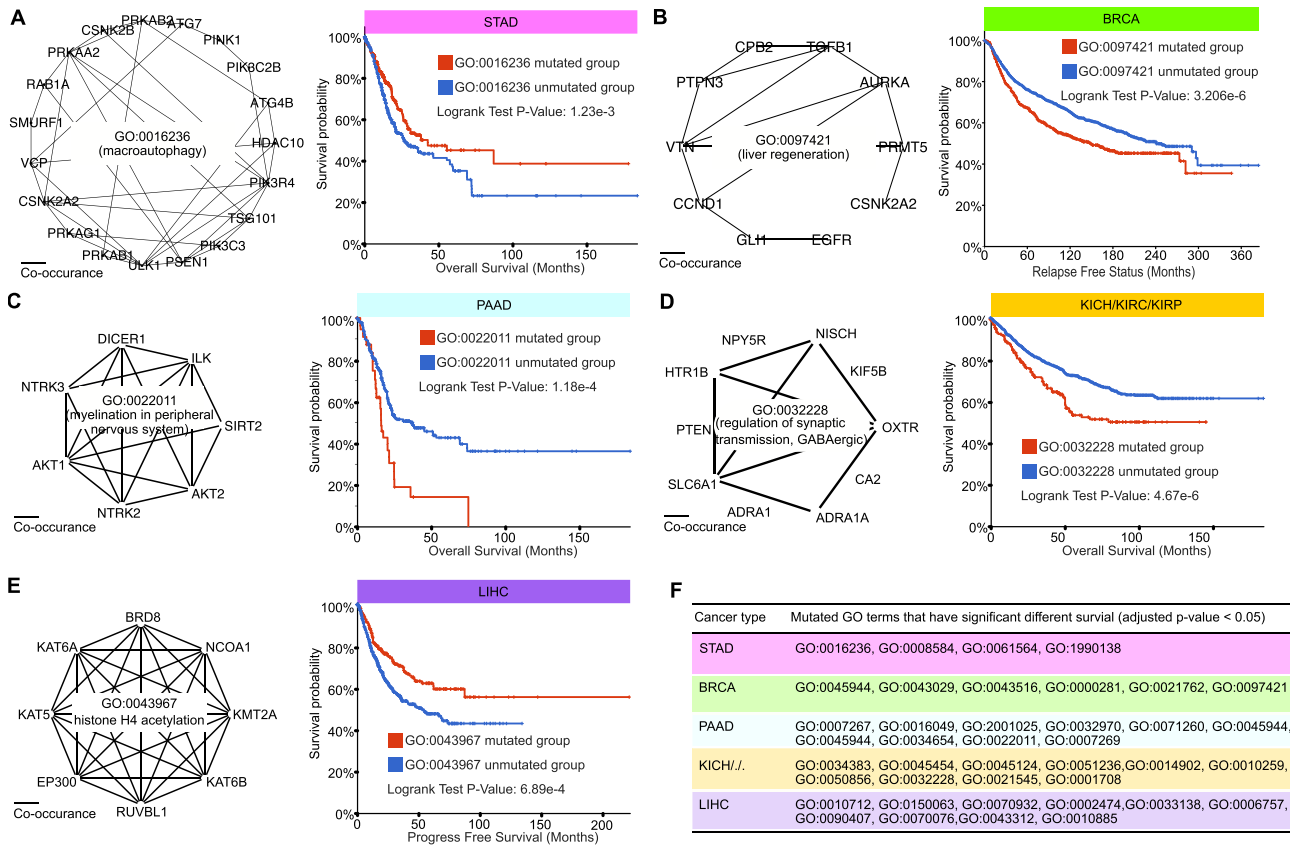


Figure 5. Examples of the GO terms identified by ParsVNN for the five cancer-specific models. (A) GO:0016236 is a GO term in the STAD-specific ParsVNN model. We study how the mutation of GO:0016236 influences the survival in patients. (B) GO:0097421 is a GO term in the BRCA-specific ParsVNN model. We study how the mutation of GO:0097421 influences the survival in patients. (C) GO:0022011 is a GO term in the PAAD-specific ParsVNN model. We study how the mutation of GO:0022011 influences the survival in patients. (D) GO:0032228 is a GO term in the KICH/KIRC/KIRP-specific ParsVNN model. We study how the mutation of GO:0032228 influences the survival in patients. (E) GO:0043967 is a GO term in the LIHC-specific ParsVNN model. We study how the mutation of GO:0043967 influences the survival in patients. (F) All GO terms in the ParsVNN models lead to significantly different survival in patients (adjusted P-value < 0.05) for five cancers.

BRCA samples (45–48). We found that all six leaf GO terms showing the mutation of the GO terms would lead to significantly different survival times (P -value <0.01). The details of these six GO terms are shown in Figure 5F and the genes annotated in those GO terms can be found in Supplementary Figure S2. In addition, in Figure 5B, we show the genes in GO:0037421 term and the corresponding survival analysis. The survival analysis shows that the mutation of GO:0037421 would lead to poor survival in patients.

Similarly, we analyzed the 12 leaf GO terms in the GO hierarchy identified by PAAD-specific ParsVNN model. For each GO term, we did survival analysis on samples with the term mutated and samples with the term unmutated using PAAD samples (45,49–56). There are 10 out of 12 leaf GO terms showing that the mutation of the GO terms would lead to significantly different survival times (P -value <0.01). The details of these 10 GO terms are shown in Figure 5F and the genes annotated in those GO terms can be found in Supplementary Figure S3. We further illustrated the genes in GO:0022011 and the co-occurrence pattern between these genes in Figure 5C. We also show the survival analysis (Figure 5C) with respect to this GO term. The survival analysis demonstrates that the mutation of GO:0022011 would lead to poor survival in patients.

In the KICH/KIRC/KIPIC-specific ParsVNN model, there are 24 leaf GO terms in the corresponding GO hierarchy. We discovered that the mutation of 10 out of 24 GO terms would lead to significantly different survival times in patients. The details of these nine GO terms are listed in Figure 5F and the genes annotated in those GO terms can be found in Supplementary Figure S4. We illustrated the genes in GO:0022011 and the co-occurrence pattern between these genes in Figure 5D. We also show the survival analysis (Figure 5D) with respect to this GO term (45,57–65). The survival analysis shows that the mutation of GO:0032228 would lead to poor survival in patients.

In the LIHC-specific ParsVNN model, there are 18 leaf GO terms in the corresponding GO hierarchy. We discovered that the mutation of 10 out of 18 GO terms would lead to significantly different survival times in patients. The details of these eight GO terms are listed in Figure 5F and the genes annotated in those GO terms can be found in Supplementary Figure S5. In Figure 5E, we listed the genes in GO:0043967 and the co-occurrence pattern between these genes and the survival analysis with respect to this GO term (45,66–71). The survival analysis shows that the mutation of GO:0043967 would benefit the survival of patients.

ParsVNN models suggest synergistic drug combination opportunities

In this section, we investigated whether we can use the subsystems identified by ParsVNN to find synergistic drug combination opportunities. Suggested by the parallel pathway inhibition theory of drug synergy (72), two drugs will be synergistic if they inhibit different pathways that regulate the same essential function. Therefore, we look for drugs that target genes in different subsystems in the BRCA ParsVNN model as shown in Figure 6A. We found that dexamethasone targets NR3C1 in subsystem 2, MK-2206 targets AKT1 in subsystem 1, and doxorubicin and etoposide

target TOPA2 in subsystem 11. These drugs suggest three different drug combinations as shown in Figure 6B. We further explored the DrugCombDB to check the drug synergy of these three different drug combinations. We listed four drug synergy scores: ZIP, BLISS, LOEWE and HSA. For these four scores, positive scores imply drug synergy and negative scores imply drug antagonism. We showed the average synergy scores in DrugCombDB in Figure 6C. We found that Comb1 has a synergistic effect on the KPL1 breast cancer cell line because all four synergy scores are positive. Similarly, Comb2 has a synergistic effect on T-47D and OCUBM breast cancer cell lines. Comb3 has a synergistic effect on KPL1, MDAMB436 and OCUBM breast cancer cell lines. This analysis indicates that we can analyze the subsystems identified by ParsVNN to find synergistic drug combinations.

ParsVNN models for other cancer types also suggest novel drug combinations, which await experimental validation. We provide those novel drug combinations in Supplementary File TS1.

DISCUSSION

Deep learning models have raised unprecedented attention in the field of bioinformatics and computational biology, due to their outperformance of traditional models extracting inherently valuable information from biological networks. VNN models have been developed (17,19,20) to overcome the issue in modern ANNs that densely entangled layers are not based on human interpretable relationships such as known biological interactions. VNNs achieve this by explicitly utilizing a biological network as its architecture and assigning both artificial neurons and the edges between neurons specific biological meanings. However, VNN models are agnostic to the biological context and have limited abilities to grasp the most important, relevant and concise architecture in regard to a particular condition of interest, by including nodes and edges that are irrelevant to the downstream prediction tasks. Specifically, the training algorithm used by VNN models cannot distinguish those irrelevant nodes and edges. After training, those irrelevant edges always have nonzero weights indicating that they are making contributions to the VNN models. For example, DrugCell uses GO hierarchy as its architecture to predict drug responses. The GO hierarchy is very general and encodes all possible prior knowledge, but it might contain irrelevant GO terms when we want to build a DrugCell model that only predicts drug responses for a specific cancer type. If we cannot distinguish the essential GO terms from those irrelevant ones, a misleading interpretation could be generated. The training data provide specific information on the prediction task; therefore, it is important to be able to utilize the training data, along with the task-agnostic VNN model, for the construction of task-specific VNN models. To address this need, we introduced here a VNN learning algorithm—ParsVNN. The unique property of ParsVNN is that starting from a task-agnostic VNN model, it utilizes training data to remove redundant nodes and edges in the task-agnostic VNN model. We demonstrated that ParsVNN outperforms the original VNN models and other state-of-the-art methods to build task-specific VNN models

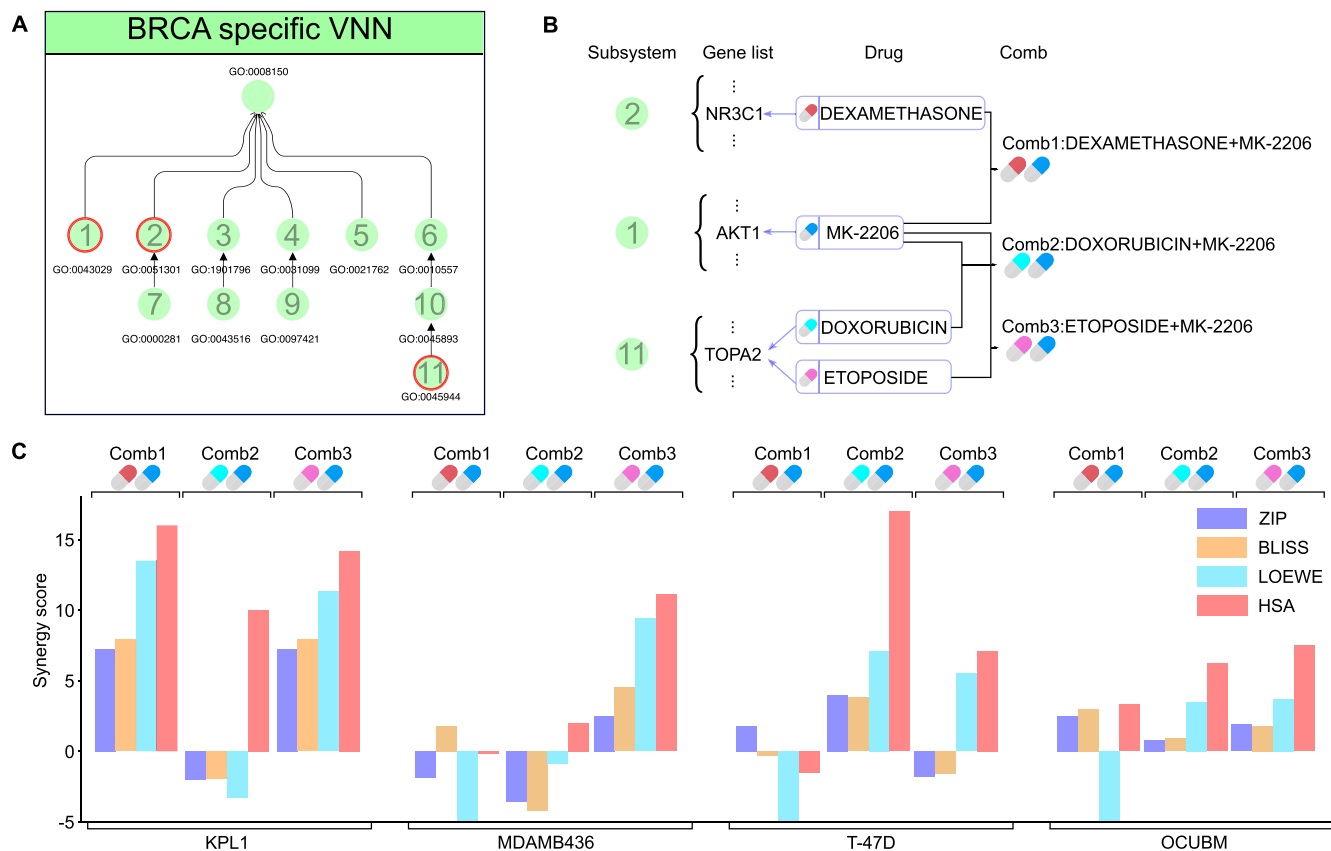


Figure 6. Drug combinations suggested by subsystems identified by ParsVNN. (A) Functional organization of subsystems in BRCA-specific ParsVNN model. (B) Genes in subsystems 1, 2 and 11 and the corresponding drugs targeting those genes. Three drug combinations are suggested by parallel pathway inhibition theory. (C) The average drug synergy scores of the three drug combinations on four different breast cancer cell lines extracted from Drug-CombDB (73). Comb1 is synergistic on KPL1. Comb2 is synergistic on T-47D and OCUBM. Comb3 is synergistic on KPL1, MDAMB436 and OCUBM.

in terms of test accuracy and interpretability (Figures 2 and 3).

We build interpretable ParsVNN models for five different cancer types and showed that the much simpler ParsVNN models contain many more cancer driver genes, co-occurrence genes and mutually exclusive genes than the competing methods (Figure 4). In addition, we discovered that most of the subsystems identified by ParsVNN have a significant impact on patient survival (Figure 5). Furthermore, we illustrated that we can analyze the subsystems identified by ParsVNN to find potential drug combinations with synergistic effects. All these aspects mentioned above demonstrated that ParsVNN models reveal insights specific to the learning task and are interpretable to bioinformaticians and clinicians alike.

Overall, our results show that ParsVNN is a very powerful method for building parsimonious VNN models that are predictive and explanatory. The explainability of our models solves one of the oldest and most challenging problems facing the biological application of deep learning. Furthermore, ParsVNN can be applied to all deep learning models that utilize biological networks in their architectures representing a broadly applicable advance in computational biology. ParsVNN can also be extended to answer a set of questions that puzzle many clinicians, biologists and researchers. One such question is how to identify unknown GO subsys-

tems, regulatory links or pathways that are not annotated in the literature or ontologies but are very informative for specific tasks. This newly discovered ontology information could be used to guide novel studies of these previously unknown mechanisms. One of our future works will be to implement a ParsVNN variant to discover interactions and links not found in the given ontology knowledge database to better guide sensible experiments and inform more precise decisions.

SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

FUNDING

This research was supported by the Indiana University Precision Health Initiative fund.

Conflict of interest statement. None declared.

REFERENCES

- Urbach,D., Lupien,M., Karagas,M.R. and Moore,J.H. (2012) Cancer heterogeneity: origins and implications for genetic association studies. *Trends Genet.*, **28**, 538–543.
- Guo,M., Peng,Y., Gao,A., Du,C. and Herman,J.G. (2019) Epigenetic heterogeneity in cancer. *Biomark. Res.*, **7**, 23.

3. Makohon-Moore, A.P., Zhang, M., Reiter, J.G., Bozic, I., Allen, B., Kundu, D., Chatterjee, K., Wong, F., Jiao, Y., Kohutek, Z.A. *et al.* (2017) Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nat. Genet.*, **49**, 358–366.
4. Reiter, J.G., Baretta, M., Gerold, J.M., Makohon-Moore, A.P., Daud, A., Iacobuzio-Donahue, C.A., Azad, N.S., Kinzler, K.W., Nowak, M.A. and Vogelstein, B. (2019) An analysis of genetic heterogeneity in untreated cancers. *Nat. Rev. Cancer*, **19**, 639–650.
5. Bedard, P.L., Hansen, A.R., Ratain, M.J. and Siu, L.L. (2013) Tumour heterogeneity in the clinic. *Nature*, **501**, 355–364.
6. Ramón Y Cajal, S., Sesé, M., Capdevila, C., Aasen, T., De Mattos-Arruda, L., Diaz-Cano, S.J., Hernández-Losa, J. and Castellví, J. (2020) Clinical implications of intratumor heterogeneity: challenges and opportunities. *J. Mol. Med.*, **98**, 161–177.
7. Dagogo-Jack, I. and Shaw, A.T. (2018) Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.*, **15**, 81–94.
8. Lim, Z.F. and Ma, P.C. (2019) Emerging insights of tumor heterogeneity and drug resistance mechanisms in lung cancer targeted therapy. *J. Hematol. Oncol.*, **12**, 134.
9. Fittall, M.W. and Van Loo, P. (2019) Translating insights into tumor evolution to clinical practice: promises and challenges. *Genome Med.*, **11**, 20.
10. Jiao, W., Atwal, G., Polak, P., Karlic, R., Cuppen, E., Danyi, A., de Ridder, J., van Herpen, C., Lolkema, M.P., Steeghs, N. *et al.* (2020) A deep learning system accurately classifies primary and metastatic cancers using passenger mutation patterns. *Nat. Commun.*, **11**, 728.
11. Sun, Y., Zhu, S., Ma, K., Liu, W., Yue, Y., Hu, G., Lu, H. and Chen, W. (2019) Identification of 12 cancer types through genome deep learning. *Sci. Rep.*, **9**, 17256.
12. Yuan, Y., Shi, Y., Li, C., Kim, J., Cai, W., Han, Z. and Feng, D.D. (2016) DeepGene: an advanced cancer type classifier based on deep learning and somatic point mutations. *BMC Bioinformatics*, **17**, 476.
13. Luo, P., Ding, Y., Lei, X. and Wu, F.X. (2019) deepDriver: predicting cancer driver genes based on somatic mutations using deep convolutional neural networks. *Front. Genet.*, **10**, 13.
14. Echle, A., Rindtorff, N.T., Brinker, T.J., Luedde, T., Pearson, A.T. and Kather, J.N. (2021) Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br. J. Cancer*, **124**, 686–696.
15. Landhuis, E. (2020) Deep learning takes on tumours. *Nature*, **580**, 551–553.
16. Kleppe, A., Skrede, O.J., De Raedt, S., Liestøl, K., Kerr, D.J. and Danielsen, H.E. (2021) Designing deep learning studies in cancer diagnostics. *Nat. Rev. Cancer*, **21**, 199–211.
17. Ma, J., Yu, M.K., Fong, S., Ono, K., Sage, E., Demchak, B., Sharan, R. and Ideker, T. (2018) Deep learning to model the hierarchical structure and function of a cell. *Nat. Methods*, **15**, 290–298.
18. Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F.C.P., Clarke, D., Gu, M., Emani, P., Yang, Y.T. *et al.* (2018) Comprehensive functional genomic resource and integrative model for the human brain. *Science*, **362**, eaat8464.
19. Kuenzi, B.M., Park, J., Fong, S.H., Sanchez, K.S., Lee, J., Kreisberg, J.F., Ma, J. and Ideker, T. (2020) Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer Cell*, **38**, 672–684.
20. Elmarakeby, H.A., Hwang, J., Liu, D., AlDubayan, S.H., Salari, K., Richter, C., Arnoff, T.E., Park, J., Hahn, W.C. and Allen, E.V. (2021) Biologically informed deep neural network for prostate cancer classification and discovery. *Nature*, **598**, 348–352.
21. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. *et al.* (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
22. Carbon, S., Douglass, E., Good, B.M., Unni, D.R., Harris, N.L., Mungall, C.J., Basu, S., Chisholm, R.L., Dodson, R.J., Hartline, E. *et al.* (2021) The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Res.*, **49**, D325–D334.
23. Seashore-Ludlow, B., Rees, M.G., Cheah, J.H., Kokol, M., Price, E.V., Coletti, M.E., Jones, V., Bodycombe, N.E., Soule, C.K., Gould, J. *et al.* (2015) Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov.*, **5**, 1210–1223.
24. Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R. *et al.* (2013) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.*, **41**, D955–D961.
25. Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Tamborero, D., Schroeder, M.P., Jene-Sanz, A., Santos, A. and Lopez-Bigas, N. (2013) IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods*, **10**, 1081–1082.
26. Bolte, J., Sabach, S. and Teboulle, M. (2014) Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.*, **146**, 459–494.
27. Wang, Y., Cho, D.Y., Lee, H., Fear, J., Oliver, B. and Przytycka, T.M. (2018) Reprogramming of regulatory network using expression uncovers sex-specific gene regulation in *Drosophila*. *Nat. Commun.*, **9**, 4061.
28. Bertsimas, D., Pauphilet, J. and Parys, B.V. (2020) Sparse regression: scalable algorithms and empirical performance. *Stat. Sci.*, **35**, 555–578.
29. Bolte, J., Sabach, S. and Teboulle, M. (2014) Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.*, **146**, 459–494.
30. Bach, F.R., Mairal, J. and Ponce, J. (2008) Convex sparse matrix factorizations. arXiv doi: <https://arxiv.org/abs/0812.1869>, 10 December 2008, preprint: not peer reviewed.
31. Rogers, D. and Hahn, M. (2010) Extended-connectivity fingerprints. *J. Chem. Inf. Model.*, **50**, 742–754.
32. Eskicak, B., McMillan, E.A., Mendiratta, S., Kollipara, R.K., Zhang, H., Humphries, C.G., Wang, C., Garcia-Rodriguez, J., Ding, M., Zaman, A. *et al.* (2017) Biomarker accessible and chemically addressable mechanistic subtypes of BRAF melanoma. *Cancer Discov.*, **7**, 832–851.
33. Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H. *et al.* (2016) A landscape of pharmacogenomic interactions in cancer. *Cell*, **166**, 740–754.
34. Potts, M.B., McMillan, E.A., Rosales, T.I., Kim, H.S., Ou, Y.H., Toombs, J.E., Brekken, R.A., Minden, M.D., MacMillan, J.B. and White, M.A. (2015) Mode of action and pharmacogenomic biomarkers for exceptional responders to didemnin B. *Nat. Chem. Biol.*, **11**, 401–408.
35. Zou, H. and Hastie, T. (2005) Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B*, **67**, 301–320.
36. authors listed, N. (2015) Correction: cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS One*, **10**, e0131499.
37. Dong, X., Huang, D., Yi, X., Zhang, S., Wang, Z., Yan, B., Chung Sham, P., Chen, K. and Jun Li, M. (2020) Diversity spectrum analysis identifies mutation-specific effects of cancer driver genes. *Commun. Biol.*, **3**, 6.
38. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E. *et al.* (2012) The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.*, **2**, 401–404.
39. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E. *et al.* (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal*, **6**, p11.
40. Guo, Y.A., Chang, M.M., Huang, W., Ooi, W.F., Xing, M., Tan, P. and Skanderup, A.J. (2018) Mutation hotspots at CTCF binding sites coupled to chromosomal instability in gastrointestinal cancers. *Nat. Commun.*, **9**, 1520.
41. Wang, K., Yuen, S.T., Xu, J., Lee, S.P., Yan, H.H., Shi, S.T., Siu, H.C., Deng, S., Chu, K.M., Law, S. *et al.* (2014) Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.*, **46**, 573–582.
42. Bass, A.J., Thorsson, V., Shmulevich, I., Reynolds, S.M., Miller, M., Bernard, B., Hinoue, T., Laird, P.W., Curtis, C., Shen, H. *et al.* (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*, **513**, 202–209.
43. Kakiuchi, M., Nishizawa, T., Ueda, H., Gotoh, K., Tanaka, A., Hayashi, A., Yamamoto, S., Tatsuno, K., Katoh, H., Watanabe, Y. *et al.* (2014) Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat. Genet.*, **46**, 583–587.
44. Wang, K., Kan, J., Yuen, S.T., Shi, S.T., Chu, K.M., Law, S., Chan, T.L., Kan, Z., Chan, A.S., Tsui, W.Y. *et al.* (2011) Exome sequencing

- identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat. Genet.*, **43**, 1219–1223.
45. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M., Chang, K. *et al.* (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*, **45**, 1113–1120.
 46. Pereira, B., Chin, S.F., Rueda, O.M., Vollen, H.K., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.J. *et al.* (2016) The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat. Commun.*, **7**, 11479.
 47. Rueda, O.M., Sammut, S.J., Seoane, J.A., Chin, S.F., Caswell-Jin, J.L., Callari, M., Batra, R., Pereira, B., Bruna, A., Ali, H.R. *et al.* (2019) Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature*, **567**, 399–404.
 48. Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y. *et al.* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, **486**, 346–352.
 49. Jiao, Y., Yonescu, R., Offerhaus, G.J., Klimstra, D.S., Maitra, A., Eshleman, J.R., Herman, J.G., Poh, W., Pelosof, L., Wolfgang, C.L. *et al.* (2014) Whole-exome sequencing of pancreatic neoplasms with acinar differentiation. *J. Pathol.*, **232**, 428–435.
 50. Wu, J., Jiao, Y., Dal Molin, M., Maitra, A., de Wilde, R.F., Wood, L.D., Eshleman, J.R., Goggins, M.G., Wolfgang, C.L., Canto, M.I. *et al.* (2011) Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. *Proc. Natl Acad. Sci. U.S.A.*, **108**, 21188–21193.
 51. Biankin, A.V., Waddell, N., Kassahn, K.S., Gingras, M.C., Muthuswamy, L.B., Johns, A.L., Miller, D.K., Wilson, P.J., Patch, A.M., Wu, J. *et al.* (2012) Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature*, **491**, 399–405.
 52. Bailey, P., Chang, D.K., Nones, K., Johns, A.L., Patch, A.M., Gingras, M.C., Miller, D.K., Christ, A.N., Bruxner, T.J., Quinn, M.C. *et al.* (2016) Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature*, **531**, 47–52.
 53. Witkiewicz, A.K., McMillan, E.A., Balaji, U., Baek, G., Lin, W.C., Mansour, J., Mollae, M., Wagner, K.U., Koduru, P., Yopp, A. *et al.* (2015) Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat. Commun.*, **6**, 6744.
 54. Cao, Y., Gao, Z., Li, L., Jiang, X., Shan, A., Cai, J., Peng, Y., Li, Y., Jiang, X., Huang, X. *et al.* (2013) Whole exome sequencing of insulinoma reveals recurrent T372R mutations in YY1. *Nat. Commun.*, **4**, 2810.
 55. Jiao, Y., Shi, C., Edil, B.H., de Wilde, R.F., Klimstra, D.S., Maitra, A., Schulick, R.D., Tang, L.H., Wolfgang, C.L., Choti, M.A. *et al.* (2011) DAXX/ATRAX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science*, **331**, 1199–1203.
 56. Scarpa, A., Chang, D.K., Nones, K., Corbo, V., Patch, A.M., Bailey, P., Lawlor, R.T., Johns, A.L., Miller, D.K., Mafficini, A. *et al.* (2017) Whole-genome landscape of pancreatic neuroendocrine tumours. *Nature*, **543**, 65–71.
 57. Miao, D., Margolis, C.A., Gao, W., Voss, M.H., Li, W., Martini, D.J., Norton, C., Bossé, D., Wankowicz, S.M., Cullen, D. *et al.* (2018) Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Science*, **359**, 801–806.
 58. Guo, G., Gui, Y., Gao, S., Tang, A., Hu, X., Huang, Y., Jia, W., Li, Z., He, M., Sun, L. *et al.* (2011) Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. *Nat. Genet.*, **44**, 17–19.
 59. Gerlinger, M., Horswell, S., Larkin, J., Rowan, A.J., Salm, M.P., Varela, I., Fisher, R., McGranahan, N., Matthews, N., Santos, C.R. *et al.* (2014) Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.*, **46**, 225–233.
 60. Creighton, C.J., Morgan, M., Gunaratne, P.H., Wheeler, D.A., Gibbs, R.A., Robertson, A., Chu, A., Beroukhi, R., Cibulskis, K., Signoretti, S. *et al.* (2013) Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*, **499**, 43–49.
 61. Sato, Y., Yoshizato, T., Shiraiishi, Y., Maekawa, S., Okuno, Y., Kamura, T., Shimamura, T., Sato-Otsubo, A., Nagae, G., Suzuki, H. *et al.* (2013) Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.*, **45**, 860–867.
 62. Davis, C.F., Ricketts, C.J., Wang, M., Yang, L., Cherniack, A.D., Shen, H., Buhay, C., Kang, H., Kim, S.C., Fahey, C.C. *et al.* (2014) The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell*, **26**, 319–330.
 63. Durinck, S., Stawiski, E.W., Pavia-Jiménez, A., Modrusan, Z., Kapur, P., Jaiswal, B.S., Zhang, N., Toffessi-Tcheuyap, V., Nguyen, T.T., Pahuja, K.B. *et al.* (2015) Spectrum of diverse genomic alterations define non-clear cell renal carcinoma subtypes. *Nat. Genet.*, **47**, 13–21.
 64. Chen, Y.B., Xu, J., Skanderup, A.J., Dong, Y., Brannon, A.R., Wang, L., Won, H.H., Wang, P.I., Nanjangud, G.J., Jungbluth, A.A. *et al.* (2016) Molecular analysis of aggressive renal cell carcinoma with unclassified histology reveals distinct subsets. *Nat. Commun.*, **7**, 13131.
 65. Chun, H.E., Lim, E.L., Heravi-Moussavi, A., Saberi, S., Mungall, K.L., Bilenky, M., Carles, A., Tse, K., Shlafman, I., Zhu, K. *et al.* (2016) Genome-wide profiles of extra-cranial malignant rhabdoid tumors reveal heterogeneity and dysregulated developmental pathways. *Cancer Cell*, **29**, 394–406.
 66. Pilati, C., Letouzé, E., Nault, J.C., Imbeaud, S., Boulai, A., Calderaro, J., Poussin, K., Franconi, A., Couchy, G., Morcrette, G. *et al.* (2014) Genomic profiling of hepatocellular adenomas reveals recurrent FRK-activating mutations and the mechanisms of malignant transformation. *Cancer Cell*, **25**, 428–441.
 67. Harding, J.J., Nandakumar, S., Armenia, J., Khalil, D.N., Albano, M., Ly, M., Shia, J., Hechtman, J.F., Kundra, R., El Dika, I. *et al.* (2019) Prospective genotyping of hepatocellular carcinoma: clinical implications of next-generation sequencing for matching patients to targeted and immune therapies. *Clin. Cancer Res.*, **25**, 2116–2126.
 68. Schulze, K., Imbeaud, S., Letouzé, E., Alexandrov, L.B., Calderaro, J., Rebouissou, S., Couchy, G., Meiller, C., Shinde, J., Soysouvanh, F. *et al.* (2015) Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.*, **47**, 505–511.
 69. Zheng, J., Sadot, E., Vigidal, J.A., Klimstra, D.S., Balachandran, V.P., Kingham, T.P., Allen, P.J., D’Angelica, M.I., DeMatteo, R.P., Jarnagin, W.R. *et al.* (2018) Characterization of hepatocellular adenoma and carcinoma using microRNA profiling and targeted gene sequencing. *PLoS One*, **13**, e0200776.
 70. Ahn, S.M., Jang, S.J., Shim, J.H., Kim, D., Hong, S.M., Sung, C.O., Baek, D., Haq, F., Ansari, A.A., Lee, S.Y. *et al.* (2014) Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. *Hepatology*, **60**, 1972–1982.
 71. Fujimoto, A., Totoki, Y., Abe, T., Boroevich, K.A., Hosoda, F., Nguyen, H.H., Aoki, M., Hosono, N., Kubo, M., Miya, F. *et al.* (2012) Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.*, **44**, 760–764.
 72. Michel, J.B., Yeh, P.J., Chait, R., Moellering, R.C. and Kishony, R. (2008) Drug interactions modulate the potential for evolution of resistance. *Proc. Natl Acad. Sci. U.S.A.*, **105**, 14918–14923.
 73. Liu, H., Zhang, W., Zou, B., Wang, J., Deng, Y. and Deng, L. (2020) DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Res.*, **48**, D871–D881.