Review

Computational Tactics for Precision Cancer Network Biology

Heewon Park [1,*] and **Satoru Miyano** [1,2]

1   M&D Data Science Center, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan
2   Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan
*   Correspondence: hwpark.dsc@tmd.ac.jp

Abstract: Network biology has garnered tremendous attention in understanding complex systems of cancer, because the mechanisms underlying cancer involve the perturbations in the specific function of molecular networks, rather than a disorder of a single gene. In this article, we review the various computational tactics for gene regulatory network analysis, focused especially on personalized anti-cancer therapy. This paper covers three major topics: (1) cell line's (or patient's) cancer characteristics specific gene regulatory network estimation, which enables us to reveal molecular interplays under varying conditions of cancer characteristics of cell lines (or patient); (2) computational approaches to interpret the multitudinous and massive networks; (3) network-based application to uncover molecular mechanisms of cancer and related marker identification. We expect that this review will help readers understand personalized computational network biology that plays a significant role in precision cancer medicine.

1. Introduction

Gene regulatory network describes functional interactions between genes, where the network is presented by a graph whose nodes present the genes, and the edges between nodes represent the regulatory interactions between genes [1,2]. Heterogeneous gene regulatory system is a useful tool to analyse and visualize biological activities and is crucial to understanding complex biological processes of cancer, because the molecular mechanisms underlying diseases reflect the perturbations in a specific function of molecules in the complex cellular network, rather than a consequence of an abnormality in a single gene [3].

The molecular interplays between genes involved in cellular processes and pathways can be represented by statistical and mathematical models. The computational strategies to estimate large-scale gene networks from gene expression levels have drawn a large amount of attention. The Gaussian graphical model (GGM), that is the probability model, has often been used to infer the conditional dependence structure of a set of genes. The GGM represents which genes (variables) predict one another and allows for sparse modeling of covariance structures, highlighting potential causal relationships between genes [4]. The Bayesian network (BN) is also a probabilistic graphical model describing a directed acyclic graph. BN has been used to uncover cancer mechanisms, i.e., unique cancer molecular mechanisms of clone cancer [5], causal networks of breast metastasis to bone, brain, or lung [6], assessing the risk of breast cancer [7], etc. Boolean networks are discrete models and one of the most widely used techniques to estimate the gene regulatory system. In the model, gene expression levels are discretized, and each gene takes on two values, i.e., if the gene expression is above a threshold value, then 1, otherwise 0, and the interactions between genes are described by standard logic (Boolean) functions [8]. Various cancer

research has been based on Boolean networks for cancer drug discovery [9], identifying lung cancer diagnostic and prognostic biomarkers [10], uncovering the mechanisms of tumorigenesis and possible treatment responses of prostate cancer [11], etc. Additionally, various computational models and strategies (e.g., differential equation-based Model, artificial neural network (ANN) approaches, correlation network, information theory, etc.) have been developed and applied to cancer research. Furthermore, the effectiveness of the networks-based analysis has been proven in various fields of research, e.g., cancer prediction, drug combinations identification, and protein-protein interaction [12–14].

Although many computational tactics for gene regulatory network estimation have been developed and numerous studies have been conducted to uncover cancer mechanisms based on the estimated gene networks, the existing studies were conducted by an averaged gene network for all cell lines. Thus, we cannot effectively identify crucial information for precision cancer medicine.

In this article, we reviewed the computational strategies for the cell line's (or patient's) cancer characteristics specific gene network analysis. Especially, we reviewed machine learning approaches for varying coefficient models, where the varying coefficients describe the strength of the interaction between genes for a specific characteristic of each cell line. That is, the model enables us to construct a gene regulatory network for a specific status related to cancer of the cell line. The cell line characteristic specific gene networks estimation provides hundreds of networks for hundreds of cell lines, where each network is given as a matrix form with about 20,000 columns for target genes, 2000 rows for regulator genes, and the elements of the matrix indicate the strength of interaction between the regulator and target genes. The analysis and interpretation of the multiple and massive networks are quite difficult tasks and have remained a serious challenge in computational biology. In this article, we also review some computational tactics for comprehensive analysis and interpretation of the large-scale networks.

The remainder of this paper is organized as follows. In the gene regulatory network estimation section, the regression framework to gene regulatory network estimation is represented. The computation tactics estimate the cell line characteristic specific gene regulatory network in the sample-specific gene network estimation section. In the section of gene network analysis in multi-dimensional cell line space, the machine learning and Artificial Intelligence (AI) approaches to comprehensive analysis of the estimated multiple and massive gene regulatory network are represented. In the Applications section, the application result of the reviewed computation strategies for network-based anti-cancer drug prediction and related markers identification is introduced. Conclusions are provided in the Discussion section.

2. Gene Regulatory Network Estimation

Suppose $X = (x_1, \ldots, x_n)^T \in \mathbb{R}^{n \times p}$ is an $n \times p$ data matrix describing the expression of p regulators that may control the transcription of ℓth target gene $y_\ell \in \mathbb{R}^n, \ell = 1, \ldots, q$. Consider the linear regression model,

$$y_\ell = \sum_{j=1}^{p} \beta_{j\ell} x_j + \epsilon_\ell, \quad \ell = 1, \ldots, q, \tag{1}$$

where $\beta_{j\ell}$ describes the effect of the jth regulator gene on the ℓth target gene, and ε_ℓ is a random error vector $\varepsilon_\ell = (\varepsilon_{\ell 1}, \ldots, \varepsilon_{\ell n})^T$ that is assumed to be independently and identically distributed with mean 0 and variance σ^2. To estimate the gene regulatory network, the following L_1-type regularization methods were used successfully,

$$L(\boldsymbol{\beta}_\ell) = \underset{\boldsymbol{\beta}_\ell}{\arg \min} \{ \frac{1}{2} \sum_{i=1}^{n} (y_{i\ell} - \sum_{j=1}^{p} \beta_{j\ell} x_{ij})^2 + P(\boldsymbol{\beta}_\ell) \}, \tag{2}$$

where

- ridge [15]: $P(\boldsymbol{\beta}_\ell) = \lambda \sum_{j=1}^{p} \beta_{j\ell}^2$
- lasso [16]: $P(\boldsymbol{\beta}_\ell) = \lambda \sum_{j=1}^{p} |\beta_{j\ell}|$
- elastic net [17]: $P(\boldsymbol{\beta}_\ell) = \lambda \sum_{j=1}^{p} \{\gamma\beta_{j\ell}^2 + (1-\gamma)|\beta_{j\ell}|\}$
- etc.

and $\lambda, \gamma > 0$ are the regularization parameters, where λ controls model complexity, and γ is a mixing parameter between the lasso and ridge penalties. The L_1-type regularization methods enable us to simultaneously identify crucial regulators and estimate their effect on a target gene. In particular, the methods effectively perform analysis of the high dimensional genomic alterations dataset.

Although the methods successfully perform edge selection and network estimation, the approaches provide an averaged network for all n cell-lines. Thus, we cannot estimate cell line (or patient) characteristic-specific models (i.e., molecular interplay). In other words, the methods are not enough to extract useful information for precision medicine.

3. Sample-Specific Gene Network Estimation

To effectively extract crucial information for precision medicine, cell line (or patient) characteristic-specific identification is a crucial issue. We reviewed computational approaches for cell-line characteristic-specific modelling, especially cell line characteristic-specific gene regulatory network estimation. The following varying coefficient model was used for cell-line characteristic-specific modelling [18],

$$y_\ell = \sum_{j=1}^{p} \beta_{j\ell}(m_\alpha) \cdot x_j + \epsilon_\ell, \quad \ell = 1, \ldots, q, \quad \alpha = 1, \ldots, n, \tag{3}$$

where $\beta_{j\ell}(m_\alpha)$ describes the effect of the jth regulator gene on the ℓth target gene in the αth target cell line. m_α is a cancer related characteristic of the αth cell lines, such as drug sensitivity and survival risk of cell lines. The model enables us to describe cell-line characteristic-$(M = m_\alpha)$ specific molecular interplay between genes, i.e., $\beta_{j\ell}(m_\alpha)$.

3.1. NetworkProfiler

The varying coefficient $\beta_{j\ell}(m_\alpha)$ describing cell-line characteristic-specific strength of the relationship between the jth regulator and the ℓth target genes in the αth cell line can be estimated by the following kernel-based L_1-type regularization method, called a NetworkProfiler [19],

$$L(\boldsymbol{\beta}_{\ell\alpha}|b_\ell) = \frac{1}{2}\sum_{i=1}^{n}\{y_{i\ell} - \sum_{j=1}^{p} \beta_{j\ell}(m_\alpha)x_{ij}\}^2 G(m_i - m_\alpha|b_\ell) + P(\boldsymbol{\beta}_{\ell\alpha}), \tag{4}$$

where

$$G(m_i - m_\alpha|b_\ell) = \exp\left\{\frac{-(m_i - m_\alpha)^2}{b_\ell}\right\}, \tag{5}$$

is a Gaussian kernel function to control the weight of cell lines when modelling the αth target cell line. The NetworkProfiler groups cell lines, according to the similarity of the specific characteristics of cell lines (i.e., modulator m_i for $i = 1, \ldots, n$), and performs modelling for α^{th} cell lines, based only on the cell-lines in the neighbourhood around the αth cell line. That is, the modelling for the αth cell line is based only on the cell lines having similar modulator characteristics to with the target sample's modulator value m_α. This implies that the NetworkProfiler can estimate cell line characteristic -specific gene regulatory networks.

The cancer-related characteristics of cell lines are not usually uniformly distributed. Figure 1 shows the anti-cancer drug sensitivity of cell lines, where the eight drugs are randomly selected from Genomics of Drug Sensitivity in Cancer (GDSC) and Cancer

Dependency Map (DepMap) projects. As shown in Figure 1, sensitivities of some anticancer drugs (characteristics of cell lines: modulator) are non-uniformly distributed, i.e., there are cell lines having rare cancer characteristics.
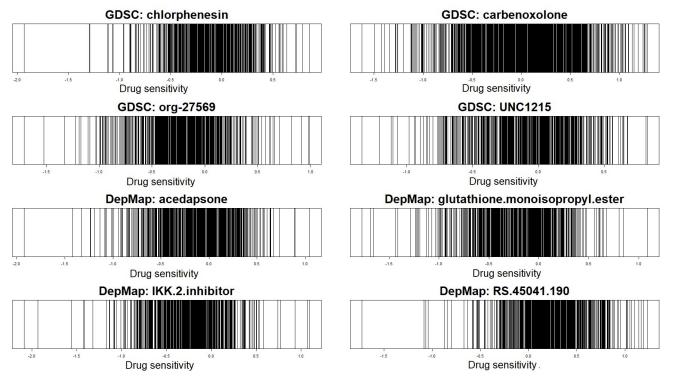

Figure 1. Drug sensitivities of GDSC and DepMap databases: each of the four drugs are randomly selected from GDSC and DepMap datsets.

Limitation: The NetworkProfiler cannot perform well when the modulator is not uniformly distributed, especially when modelling the target cell line with a rare characteristic located in a sparse region of its distribution, because the method is based on the constant bandwidth (b_ℓ) of Gaussian kernel function. In the NetworkProfiler, the bandwidth specifies the length-scale of the kernel function and controls the weights of cell lines. It implies that the NetworkProfiler based on the constant bandwidth performs cell line characteristic-specific modelling without consideration of the distribution of the modulator and location of the modulator value m_α of the target sample in the distribution. Thus, the NetworkProfiler imposes a small amount of weight to almost all samples for modelling a target sample in a sparse region. Figure 2 shows the values of the Gaussian kernel function (i.e., weight for cell lines) with a constant bandwidth b_ℓ for a target sample in both sparse and dense regions, where y-axis and x-axis indicate weights and modulator values of cell-lines, respectively. As shown in Figure 2, the Gaussian kernel function based on the constant bandwidth imposes the non-zero weight on only a few samples for the modelling of the target sample in a sparse region. It leads to extremely high dimensional data situations; thus, gene regulatory network estimation (i.e., edges selection and edge size estimation) cannot be appropriately performed.
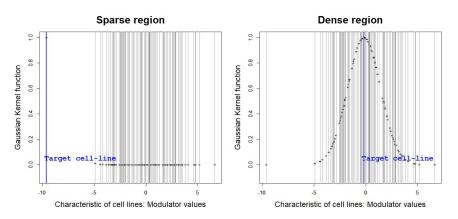
Figure 2. Gaussian kernel function to impose weight on cell lines, where *y*-axis and *x*-axis indicate weight and modulator values of cell lines.

3.2. Adaptive NetworkProfiler

To settle on the issue, Park et al. [20] developed a novel strategy, called an adaptive NetworkProfiler, based on the adaptive bandwidth of the Gaussian function. The adaptive NetworkProfiler computes the weight of cell lines by using the adaptive Gaussian kernel function, where the bandwidth is based on the *k*-nearest neighbour (KNN) rule, called an adaptive bandwidth [21]. The adaptive bandwidth for an αth target cell line is based on the Euclidean distance between the modulator value of αth cell line (m_α) and its kth nearest neighbour. By using not a constant but the adaptive bandwidth based on Euclidean distance, the KNN-Gaussian kernel function has a relatively wide width of the kernel for modelling a target sample, having a rare modulator value located in the sparse region, because the kth nearest neighbourhood of αth target cell line is also far from the m_α. Thus, the KNN-Gaussian kernel function can overcome the drawback of the ordinary NetworkProfiler for modelling a target sample in the sparse region.

The adaptive NetworkProfiler was developed based on the adaptive kennel function, with an additional parameter incorporating dispersion of modulators (i.e., range of a modulator: $r(M)$) as follows,

$$L(\boldsymbol{\beta}_{\ell\alpha}|b_\ell^{\text{KNN}}, r(M)) = \frac{1}{2}\sum_{i=1}^{n}\{y_{i\ell} - \sum_{j=1}^{p}\beta_{j\ell}(m_\alpha)x_{ij}\}^2 K(m_i - m_\alpha|b_{\ell\alpha,r(M)}^{\text{KNN}}) + P(\boldsymbol{\beta}_{\ell\alpha}), \quad (6)$$

where

$$K(m_i - m_\alpha|b_{\ell\alpha}^{\text{KNN}}, r(M)) = \exp\left(\frac{-(m_i - m_\alpha)^2}{b_{\ell\alpha}^{\text{KNN}} \cdot r(M)}\right), \quad (7)$$

and $b_{\ell\alpha}^{\text{KNN}}$ is the Euclidean distance between m_α, with its kth nearest neighbour $m_\alpha^{k^{th}}$,

$$b_{\ell\alpha}^{\text{KNN}} = \sqrt{(m_\alpha - m_\alpha^{k^{th}})^2} \quad \text{for} \quad \alpha = 1, 2, \ldots, n, \quad (8)$$

and $r(M)$ is the hyperparameter, incorporating dispersion of the modulator *M*.

The adaptive bandwidth in the Gaussian kernel function with the additional parameter incorporates distribution of cell line characteristics (*M*) and location of the characteristic value (m_α) in their distribution. Thus, the adaptive NetworkProfiler can overcome only a small number of samples that have non-zero weight and can effectively perform cell line characteristic-specific gene network estimation for modelling the target sample in not only dense regions but also sparse regions.

Limitation: The NetworkProfiler and Adaptive NetworkProfiler construct the cancer characteristics-specific gene network based on a specific cancer characteristic. That is, the methods consider a characteristic and measure similarity of

cell lines in one-dimensional cell line characteristic space based only on one characteristic. Thus, the cancer characteristic-specific gene networks estimated by the methods cannot described gene regulatory system under varying conditions of various cancer characteristics because the methods are based on a characteristic.

3.3. Gene Network Analysis in Multi-Dimensional Cell Line Space

In order to incorporate various cancer-related characteristics of cell lines and extract more precise cell-line specific molecular interplays, the cell line characteristic specific gene network estimation is extended to the multi-dimensional cell line space [22].

For h characteristics of cell line $M = (m_1, \ldots, m_h) \in \mathbb{R}^{n \times h}$, the varying coefficient model in (3) is given as follows,

$$y_\ell = \sum_{j=1}^{p} \beta_{j\ell}(m_\alpha) \cdot x_j + \epsilon_\ell, \quad \ell = 1, \ldots, q, \quad \alpha = 1, \ldots, n, \tag{9}$$

where $m_\alpha = (m_{\alpha 1}, \ldots m_{\alpha h})$. In the h-dimensional cell line space, the similarity between cell lines is measured by the following multivariate Gaussian kernel function,

$$K(m_i - m_\alpha | H_\ell) = |H_\ell|^{-1/2} \exp\left\{ -\frac{1}{2}(m_i - m_\alpha)^T H_\ell^{-1}(m_i - m_\alpha) \right\} \tag{10}$$

where H_ℓ is the bandwidth matrix (e.g., covariance matrix). Then, the multi-dimensional cell line characteristic specific gene network is estimated by the following multivariate kernel-based L_1-type regularization method,

$$L(\beta_{\ell\alpha} | H_\ell) = \frac{1}{2} \sum_{i=1}^{n} \{ y_{i\ell} - \sum_{j=1}^{p} \beta_{j\ell}(m_\alpha) x_{ij} \}^2 K(m_i - m_\alpha | H_\ell) + P(\beta_{\ell\alpha}). \tag{11}$$

The multi-dimensional cell line characteristic specific analysis enables us to extract more precise characterization of cell-lines, and thus we can effectively estimate precision cancer gene regulatory networks.

Limitation: The precision cancer gene networks estimation provides hundreds of matrices with more than 2000 rows for regulator genes and more than 10,000 columns for target genes. Although various computational tactics have been developed and successfully applied to gene regulatory network estimation, the interpretation of the large-scale gene networks remains a challenge. The existing studies on the cell line characteristic-specific gene networks focused only on the known markers and then interpreted the massive networks based on the neighbourhoods of the known markers, i.e., only narrow interpretation was performed. However, comprehensive analysis of the multiple massive networks is essential to understand the complex mechanism of cancer. The interpretation of the multi-layer massive network was the bottle network of the existing studies on the precision cancer gene networks analysis.

4. Interpretation of the Multi-Layer Massive Networks

In this section, we review computational strategies to interpret the multiple and massive gene regulatory networks.

4.1. Network Constrained Sparse Common Component Analysis (NetSCCA)

Park et al. [22] considered common structure identification of the multiple matrix datasets to interpret multilayer massive networks. The cell line specific gene regulatory

system can be described by the following regulatory effect of the jth regulator gene on the ℓth target gene in the αth cell line [19,22],

$$r_{\alpha l j} = \hat{\beta}_{\ell j}(m_\alpha) \cdot x_{\alpha j}, \quad \text{for} \quad j = 1, \dots, p, \tag{12}$$

where $x_{\alpha j}$ is an expression level value of the jth gene in the αth cell line. For the ℓth target gene, a matrix for the regulatory effect of p regulators is given as $\boldsymbol{R}_\ell = (\boldsymbol{r}_{1\ell}, \dots, \boldsymbol{r}_{n\ell})^T \in \mathbb{R}^{n \times p}$, where $\boldsymbol{r}_{\alpha\ell} = (r_{\alpha\ell 1}, \dots, r_{\alpha\ell p})^T$.

To interpret the large-scale gene regulatory networks and identify crucial biomarkers that play a key role in cancer-related mechanism of interest, the network-constrained sparse common component analysis (NetSCCA) was developed. The crucial common component of multiple datasets $(\boldsymbol{R}_\ell, \ell = 1, \dots, q)$ can be estimated by [23],

$$\arg \min_{\boldsymbol{A}} \{ \sum_{\ell=1}^{q} \| \boldsymbol{R}_\ell - \boldsymbol{R}_\ell \boldsymbol{A} \boldsymbol{A}^T \|_F^2 \}, \tag{13}$$

$$\text{subject to} \quad \boldsymbol{A}^T \boldsymbol{A} = \boldsymbol{I}_K.$$

As show in (13), the common component analysis of q datasets can be considered as a principal component analysis (PCA) of q datasets. That is, if there is only one dataset $\boldsymbol{R}_1$, then the model becomes a standard PCA. Wang et al. [23] showed that the common loading matrix $\boldsymbol{A}$ in (13) can be optimized as the solution to the following problem,

$$\arg \max_{\boldsymbol{A}} \text{Tr}(\boldsymbol{A}^T \boldsymbol{G} \boldsymbol{A}), \tag{14}$$

$$\text{subject to} \quad \boldsymbol{A}^T \boldsymbol{A} = \boldsymbol{I}_K.$$

where $\boldsymbol{G} = \sum_{\ell=1}^{q} \boldsymbol{R}_\ell^T \boldsymbol{R}_\ell$. It implies that the common loading matrix $\boldsymbol{A}$ can be estimated by the standard PCA problem.

$$\arg \min_{\boldsymbol{A}} \{ \| \boldsymbol{Q} - \boldsymbol{Q} \boldsymbol{A} \boldsymbol{A}^T \|_F^2 \}, \tag{15}$$

$$\text{subject to} \quad \boldsymbol{A}^T \boldsymbol{A} = \boldsymbol{I}_K,$$

where $\boldsymbol{Q}$ is the square root of $\boldsymbol{G}$, i.e., $\boldsymbol{Q}^T \boldsymbol{Q} = \boldsymbol{G}$. The common component analysis enables us to estimate the common subspace of the multiple massive networks (i.e., $\boldsymbol{R}_\ell, \ell = 1, \dots, q$), and extract the crucial common component of the datasets.

The common component estimation in (13) provides a fully dense projection matrix $\boldsymbol{A}$. That is, the common component is estimated by a linear combination of all features. It not only leads to difficult to interpret estimated common components but also erroneous estimation results, because the common component analysis is based on crucial and noisy features. To settle the issue, a sparse learning-based strategy was proposed and developed to achieve better biological interpretability, called a NetSCCA [22]. The NetSCCA estimates the projection matrix $\boldsymbol{A}$ based only on crucial features without disturbance of noisy features by using sparse learning and incorporates network biology knowledge that the genes with similar molecular interactions may have similar biological function in the common component estimation.

The NetSCCA measures the similarity between genes on networks by using the following jaccard similarity [24]:

$$W_{j,s} = \frac{|N_j \cap N_s|}{|N_j \cup N_s|} \tag{16}$$

where N_j is the set of nodes that are directly connected to the jth gene via an edge in at least one cell line. Then, the similarity between genes $W_{j,s}$ is incorporated into the sparse common loading matrix (A) estimation as follows,

$$\arg \min_{\Theta, A} \{ \sum_{\ell=1}^{q} \| Q - Q\Theta A^T \|_F^2 \} + \lambda_1 \sum_{k=1}^{K} \| \theta_k \|_1 + \lambda_2 \sum_{k=1}^{K} \sum_{j<s} (\theta_{j,k} - \theta_{s,k})^2 W_{j,s}, \qquad (17)$$

where $\theta_k \propto A_k$ is the p-dimensional vector and $\Theta = (\theta_1, \ldots, \theta_k)$. The last term (penalty term) of (17) enables us to locally smooth the coefficients and encourage the simultaneous selection of related genes. In other words, a large amount of weight is imposed on the coefficients of the two genes with many common interactions, and it encourages similarity in their coefficients of common structure estimation. Thus, the NetSCCA can provide biologically interpretable results of the common component analysis of the multiple networks.

The NetSCCA algorithm is given in Algorithm 1.

Algorithm 1 NetSCCA: Network constrained sparse common component analysis.

1: Compute jaccard similarity: W.
2: For q target genes, compute the regulator effect matrices as R_ℓ for $\ell = 1, \ldots, q$ and $G = \sum_{\ell=1}^{q} R_\ell^T R_\ell$.
3: For the square root of G (i.e., $Q^T Q = G$), compute sparse common loadings of q regulator effect matrices $R_\ell, \ell = 1, .., q$.
   3.1: Start A at $V = [V_1, V_2, \ldots, V_K]$, which is the loading matrix from ordinary PCA of Q.
   3.2: Given a fixed $A = [a_1, a_2, \ldots, a_K]$, solving the following problem,

$$\hat{\theta}_k \arg \min_{\theta_k} \{ \| z_k - Q\theta_k \|^2 \} + \lambda_1 \| \theta_k \|_1 + \lambda_2 \sum_{j<s} (\theta_{j,k} - \theta_{s,k})^2 W_{j,s} \quad k = 1, 2, \ldots, K,$$

   where $z_k = Qa_k$. Update $\hat{\Theta} = [\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K]$.
   3.3: For a fixed $\hat{\Theta}$, perform the singular value decomposition of $Q^T Q\hat{\Theta} = U\Gamma V^T$ and update $\hat{A} = UV^T$ (see Zou et al. [25]).
   3.4: Repeat Steps 3.2–3.3, until convergence.
4: Sparse common loading is given by $\frac{\hat{\theta}_k}{\| \hat{\theta}_k \|}$ for $k = 1, \ldots, K$.

Limitation: As pointed out by existing studies on network-based regularization [26,27], the network-constrained regularisation cannot perform well when the connected genes have opposite signs of coefficients. The limitation of the NetSCCA can be overcome by use of the advanced network-constrained regularization methods that incorporate signs of the regression coefficients [27].

4.2. Explainable AI for Gene Network-Based Prediction (Xprediction)

In this section, we review an explainable AI approach for the network-based prediction, called Xprediction [28]. Although the machine learning-based AI approaches provide effective prediction results, most of the existing approaches were developed focusing only on mathematical/statistical accuracy. Thus, the existing AI methodologies cannot explain their decision rules (i.e., the existing AI cannot explain how and why a decision has been made, causing the black-box problem). However, the interpretability and explainability are essential for use of AI strategies in various fields of research, especially medical science.

Xprediction achieves not only prediction accuracy but also interpretability of deep learning-based AI. The method is based on the widely used machine learning and deep learning approaches, e.g., the kernel support vector machine, random forest and deep neural network for prediction models, and describes the cruciality of input on output by comparison with the results of the model without the input. That is, Xprediction constructs

a model by removing a feature (i.e., by removing a molecular interaction between ℓ^{th} target and j^{th} regulator genes) individually and performing a prediction, and the prediction is iterated based on the randomly constructed cross-validation datasets. The cruciality of each molecular interaction was measured by comparing with the prediction accuracy based on all molecular interactions $\text{Acc}(\hat{\boldsymbol{y}})$.

The significance of each molecular interaction is computed by the *t*-test between prediction accuracies between models with and without the edge (i.e., interaction). Let N be a number of iterations for computing prediction accuracy from the randomly constructed cross-validation dataset, then $\overline{\text{Acc}(\hat{\boldsymbol{y}})}$ and $s_{\hat{\boldsymbol{y}}}$ are mean and standard deviation of the prediction of accuracies in N iterations, respectively. In the model without (l, j) interaction, corresponding notations are given $N^{(l,j)}$, $\overline{\text{Acc}(\hat{\boldsymbol{y}}^{(l,j)})}$ and $s_{(\hat{\boldsymbol{y}}^{(l,j)})}$, respectively. We performed the following *t*-test,

$$T_{\ell j} = \frac{\overline{\text{Acc}(\hat{\boldsymbol{y}})} - \overline{\text{Acc}(\hat{\boldsymbol{y}}^{(l,j)})}}{s_p \sqrt{\frac{1}{N} + \frac{1}{N^{(l,j)}}}} \tag{18}$$

where $s_p = \sqrt{\frac{s_{\hat{\boldsymbol{y}}}(N-1) + s_{\hat{\boldsymbol{y}}^{(l,j)}}(N^{(l,j)}-1)}{N + N^{(l,j)} - 2}}$. Then, the cruciality of $(l, j)^{th}$ interaction $I_{\ell j}$) on the prediction result was measured by the p value of the *t*-test. The algorithm of Xprediction is given in Algorithm 2.

Algorithm 2 Xprediction: explainable prediction.

1: Construct prediction models based on the kernel support vector machine (kSVM), Random Forest (RF), and Neural Network (NN).
2: Compute prediction accuracies based on k-fold cross-validation (CV). The average of the prediction accuracies of k validation sets was given as: $\text{Acc}(\hat{\boldsymbol{y}})$.
3: Step 2 is iterated N times for randomly constructed k-fold CV datasets.
4: If $l \leq q$, then
5: 　If $j \leq p$, then
6: 　　Delete (l, j) elements from regulatory effect matrices: $\boldsymbol{R}_\ell, \ell = 1, \ldots, q$
7: 　　Compuate prediction accuracy of the model without (l, j) elements: $\text{Acc}(\hat{\boldsymbol{y}}^{(l,j)})$.
8: 　　Step 7 is iterated $N^{(l,j)}$ times for randomly constructed k-fold CV datasets.
9: Perform *t*-test between $\text{Acc}(\hat{\boldsymbol{y}})$ and $\text{Acc}(\hat{\boldsymbol{y}}^{(l,j)})$ obtained from N and $N^{(l,j)}$ iterations and compute p value.
10: Cruciality of molecular interplays for AI-based prediction results are measured by on p value of the *t*-test.

Limitation:　The Xprediction constructs $(q \times p) + 1$ prediction models, because prediction accuracies of the model based on the regulatory effect without the (i, j) element should be compared with the model based on the regulatory effect with all elements. This leads to a great amount of computation. The computational complexity is one of limitations of the Xprediction.

5. Applications

In this section, we introduce an application of the introduced computational strategy to precision cancer network analysis. We consider the application of the explainable AI, Xprediction, to identify anti-cancer drug markers. The drug sensitivity data (i.e., primary-screen-replicate-collapsed-logfold-change) and RNA-expression levels of genes are obtained from the CCLE dataset (https://depmap.org/portal/, accessed on 4 August 2022). For expression levels of genes, we extracted 1922 genes that had the highest 10% variances in cell lines. We focused on anti-cancer drugs, capecitabine, and oxaliplatin, which are used in a chemotherapy combination known as XELOX or CAPEOX. The XELOX is used to colorectal and gastric cancer [29–31].
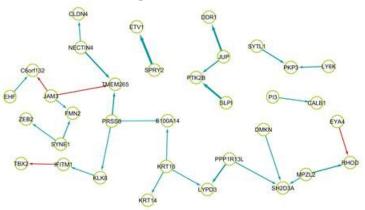
We first estimated capecitabine's sensitivity specific gene networks by using the NetworkProfiler. We then defined oxaliplatin sensitive and resistant cell lines based on fifth (5P) and ninety-fifth (95P) percentiles of the drug sensitivity (DS) values, i.e., sensitive cells: DS < 5P and resistant cells: DS > 95P. We constructed a prediction model based on the deep learning approach (i.e., deep neural network) to predict the sensitivity of the oxaliplatin. In our analysis, a two hidden layered, fully-connected feed-forward neural network was used. The *ReLU* activation function was used on the hidden layers, and the sigmoid function was used on the output layer. We randomly split the dataset 10-fold and evaluated the prediction accuracy based on the 10-fold cross-validation, i.e., prediction accuracy was given as an average of prediction accuracies of 10 test sets. By using Xprediction, crucial molecular interactions to explain sensitivities of the oxaliplatin were identified based on p value<0.05. Table 1 shows the identified crucial interactions and corresponding p value.

Table 1. Crucial molecular interplays to explain oxaliplatin sensitivity, where $X \rightarrow Y$ indicates interaction from regulator gene X to target gene Y.

Interaction	p Value	Interaction	p Value
MPZL2→SH2D3A	0.003	TP63→ITGB4	0.039
JUP→DDR1	0.008	EHF→C6orf132	0.040
DMKN→SH2D3A	0.018	PPP1R13L→LYPD3	0.040
DMKN→MPP1	0.019	IFITM1→TBX2	0.040
KRT16→S100A14	0.019	JAM3→FMN2	0.042
PRSS8→TMEM265	0.029	S100A7→KRT14	0.044
SLPI→PTK2B	0.032	KLK8→IFITM1	0.046
PI3→CALB1	0.033	EYA4→RHOD	0.046
SPRY2→ETV1	0.033	SYNE1→ZEB2	0.048
LY6K→PKP3	0.034	NECTIN4→CLDN4	0.049
SYTL1→KLK8	0.035		

Figure 3 shows gene regulatory networks consisting of the identified crucial molecular interplays to oxaliplatin sensitivity prediction, where the top and bottom indicates the networks in drug sensitive and resistance cell lines, respectively. The edge sizes represent the median of strengths of interactions between genes in drug-sensitive cell lines and -resistant cell lines, respectively.

As shown in Figure 3, drug-sensitive and-resistant cell lines show different gene regulatory systems of the identified markers. The interplay of SYNE1→ IFITM1 can be considered as a oxaliplatin-resistant specific gene regulatory system. The interplays of SPRY2→ ETV1 and SLPI → PTK1B become weaker from sensitive to resistant cell lines. It was uncovered that high expression levels of the identified drug resistant markers SYNE1 and IFITM1 are associated with poorer chemotherapy efficacy of gastric cancer and resistance to endocrine therapy and chemotherapy [32,33]. The existing studies support our results that the high activities of SYNE1 and IFITM1 are characteristics of capecitabine-resistant cell lines. On the other hand, it was demonstrated that the high expression levels of the SPRY2 are associated with chemotherapy-sensitive cell line MEK inhibitors, BRAF inhibitor-resistant cells, and ovarian cancer cells [34–36]. The results of the literature are consistent with our result that the activity of SPRY2 is a signature of capecitabine-sensitive cell lines. This implies that our gene network analysis results are strongly supported by existing literatures.

Gene network in Capecitabine sensitive cell lines

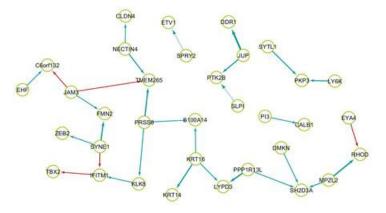

Gene network in Capecitabine resistant cell lines

Figure 3. Gene regulatory networks of the crucial molecular interplays to oxaliplatin sensitivity prediction. Color of edge indicate negative (red) and positive (blue) effects of regular genes on their target genes.

Table 2 shows that the genes consisted of the crucial interplays, related anti-cancer drugs, and cancer, where the column "Resistant" indicates that the gene was identified as a drug-resistant marker in existing studies. It can be seen from Table 2 that more than half of the identified genes are confirmed as a therapeutic target for not only XELOX (i.e., Oxaliplatin and Capecitabine) but also various anti-cancer drugs (e.g., 5-FU, cisplatin, Paclitaxel, etc.). Furthermore, the cancer-related mechanism of the genes has been verified in the literature. Although the mechanism of some genes has not yet been uncovered, it can be considered through our results and literatures that not just a single gene but the identified molecular interplays may be crucial to understanding the mechanism of anti-cancer drug resistance of cell lines.

We suggest though the application results of precision cancer network analysis and literature that molecular interplays between "SYNE1 and IFITM1" may lead to capecitabine resistance of cancer cell lines, while weakening the molecular regulatory interactions between "SPRY2 and ETV1" and "SLPI and PTK1B" induce capecitabine-resistance in cell lines.

Table 2. Identified markers and their evidence.

Genes	RG/TG	Drugs	Cancer	Resistant	Evidences
C6orf132	TG	-	-		
CALB1	TG	-	-		
CLDN4	TG	5-FU, cisplatin, Paclitaxel, cDDP	PDC, GS, CRC	*	[37–41]
DDR1	TG	Oxaliplatin	GS, CRC		[42–44]
DMKN	RG	-	CRC		[45]
EHF	RG	-	-		
ETV1	TG	oxaliplatin, 5-FU	HCC, GS, CRC	*	[46–48]
EYA4	RG	-	-		-
FMN2	TG	-	-		-
IFITM1	RG,TG	-	GS, CRC, EAC, GBC		[49–51]
ITGB4	TG	cisplatin, erlotinib, 5-Fu	LC	*	[52–54]
JAM3	RG	-	LIC		[55]
JUP	RG	-	-		-
KLK8	RG,TG	oxaliplatin	CRC, PC	*	[56,57]
KRT14	TG	Erlotinib	LC, BC		[58,59]
KRT16	RG	Erlotinib	BC		[59]
LY6K	RG	-	GC, BC	*	[60–62]
LYPD3	TG	-	AML		[63]
MPZL2	RG	-	-		-
NECTIN4	RG	5-FU, Enfortumab Vedotin	CC, BC, GC, LC	*	[64,65]
PI3	RG	-	-		-
PKP3	TG	5-FU, leucovorin, oxaliplatin	CRC		[66]
PPP1R13L	RG	5-FU	GS	*	[67]
PRSS8	RG	-	-		-
PTK2B	TG	Midostaurin, gilteritinib/defactinib, TKI	AML	*	[68]
RHOD	TG	-	-		-
S100A14	TG	-	GS		[69]
S100A7	RG	-	-		-
SH2D3A	TG	-	-		-
SLPI	RG	-	-		-
SPRY2	RG	5-FU	CC		[70]
SYNE1	RG	-	GC		[32]
SYTL1	RG	-	-		-
TBX2	TG	platinum-ased chemotherapy	OC, CRC		[71,72]
TMEM265	TG	-	-		-
TP63	RG	apatinib and capecitabine	BC		[73]
ZEB2	TG	oxaliplatin and 5-FU, cisplatin, trastuzumab	CRC, GC	*	[74–76]

AML: Acute myelogenous leukaemia; BC: breast cancer; CC: colon cancer; CRC: colorectal cancer; EAC: oesophageal adenocarcinoma; GBC: gallbladder cancer; GS: gastric cancer; HCC: hepatocellular carcinoma; LC: lung cancer; OC: ovarian cancer; PC: pancreatic cancer; PDC: pancreatic ductal carcinomas; cDDP: cis-diamminedichloroplatinum. The * indicates that the gene was identified as a drug resistant marker in the existing studies.

6. Discussion

In this article, we reviewed computational tactics for precision cancer network analysis. Although many studies have been conducted to develop computational approaches to gene regulatory network analysis and the gene network-based analysis has been applied to cancer research, the existing studies focused on an averaged gene network for all cell lines. Thus, we cannot extract crucial information for precision cancer research. In this article, we have focused on cancer characteristic-specific gene networks and reviewed the computational strategies for cell line specific modelling to identified cancer characteristic-specific molecular interplays. We also reviewed the studies on analysis and interpretation

of the estimated multiple and massive gene regulatory networks. Finally, we introduced the application results of the introduced computational tactics to anti-cancer drug sensitivity-specific gene network analysis. The application section described cell line's characteristic-(drug sensitivity) specific gene regulatory network analysis. Our analysis can be easily extended to patient's characteristic-specific gene network analysis by using expression levels and drug sensitivities summarized in each patient. We expect that the results of a cancer patient's characteristic-specific gene network analysis provides crucial evidence for precision medicine.

Although we have reviewed some computational tactics for interpretation of multiple and massive gene networks, from cell line characteristic-specific gene network estimation to computational network biology, interpretation and analysis of the large-scale gene networks is still in its infancy. Thus, researchers in various fields of research are faced with a challenge to interpret the estimated large gene networks. Explainable machine learning and, more specifically, interpretable artificial intelligence will be a key tool to overcome this bottleneck in the near future.

Author Contributions: H.P. performed data analysis for the Applications section and drafted the manuscript. S.M. supervised the works. All authors have read and approved the final version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in the Application section are from the Dependency Map (DepMap) projects (https://depmap.org/portal/, accessed on 4 August 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ristevski, B. Overview of Computational Approaches for Inference of MicroRNA-Mediated and Gene Regulatory Networks. *Advan. Comp.* **2015**, *97*, 111–145.
2. Aittokallio, T.; Schwikowski, B. Graph-based methods for analysing networks in cell biology. *Brief Bioinform.* **2006**, *7*, 243–255. [CrossRef] [PubMed]
3. Ahmed, K.; Park, S.; Jiang, Q.; Yeu, Y.; Hwang, T.; Zhang, W. Network-based drug sensitivity prediction. *BMC Med. Genom.* **2020**, *13* (Suppl. 11), 193. [CrossRef] [PubMed]
4. Epskamp, S.; Waldorp, L.; Mottus, R.; Borsboom, D. The Gaussian Graphical Model in Cross-Sectional and Time-Series Data. *Multivariate Behav. Res.* **2018**, *53*, 453–480. [CrossRef] [PubMed]
5. Liu, E.; Li, J.; Kinnebrew, G.; Zhang, P.; Zhang, Y.; Cheng, L.; Li, L. A Fast and Furious Bayesian Network and Its Application of Identifying Colon Cancer to Liver Metastasis Gene Regulatory Networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *18*, 1325–1335. [CrossRef] [PubMed]
6. Park, S.; Hwang, K.; Chung, C.; Roy, D.; Yoo, C. Causal Bayesian gene networks associated with bone, brain and lung metastasis of breast cancer. *Clin. Exp. Metastasis* **2020**, *37*, 657–674. [CrossRef]
7. Huang, Y.; Zheng, C.; Zhang, X.; Cheng, Z.; Yang, Z.; Hao, Y.; Shen, J. The Usefulness of Bayesian Network in Assessing the Risk of Triple-Negative Breast Cancer. *Acad. Radiol.* **2020**, *37*, 282–291. [CrossRef]
8. Xiao, Y. A tutorial on analysis and simulation of boolean gene regulatory network models. *Curr. Genom.* **2009**, *10*, 511–525. [CrossRef]
9. Biane, C.; Delaplace, F. Causal Reasoning on Boolean Control Networks Based on Abduction: Theory and Application to Cancer Drug Discovery. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *16*, 574–1585. [CrossRef]
10. Guo, N.; Wan, Y. Network-based identification of biomarkers coexpressed with multiple pathways. *Cancer Inform.* **2014**, *16*, 37–47. [CrossRef]

11. Montagud, A.; Beal, J.; Tobalina, L.; Traynard, P.; Subramanian, V.; Szalai, B.; Alfoldi, R.; Puskas, L.; Valencia, A.; Barillot, E.; et al. Patient-specific Boolean models of signalling networks guide personalised treatments. *eLife* **2022**, *11*, e72626. [CrossRef] [PubMed]

12. Daoud, M.; Mayo, M.; A survey of neural network-based cancer prediction models from microarray data. *Artif. Intell. Med.* **2019**, *97*, 204–214. [CrossRef] [PubMed]

13. Cheng, F.; Kovacs, I.; Barabasi, A. Network-based prediction of drug combinations. *Nat. Comm.* **2019**, *10*, 1197. [CrossRef] [PubMed]

14. Fout, A.; Byrd, J.; Shariat, B.; Ben-Hur, A. Protein interface prediction using graph convolutional networks. In Proceedings of the NIPS'17: 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6533–6542.

15. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Techonometrics* **1970**, *12*, 55–67. [CrossRef]

16. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [CrossRef]

17. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 301–320. [CrossRef]

18. Hastie, T.; Tibshirani, R. Varying-coefficient models. *J. R. Stat. Soc. Ser. B* **1993**, *55*, 757–796. [CrossRef]

19. Shimamura, T.; Imoto, S.; Shimada, Y.; Hosono, Y.; Niida, A.; Nagasaki, M.; Yamaguchi, R.; Takahashi, T.; Miyano, S. A novel network profiling analysis reveals system changes in epithelial-mesenchymal transition. *PLoS ONE* **2011**, *6*, e20804. [CrossRef]

20. Park, H.; Shimamura, T.; Imoto, S.; Miyano S. Adaptive NetworkProfiler for Identifying Cancer Characteristic-Specific Gene Regulatory Networks. *J. Comput. Biol.* **2018**, *25*, 130–145. [CrossRef]

21. Terrell, G.; Scott, D. Variable kernel density estimation. *Ann. Stat.* **1992**, *20*, 1236–1265. [CrossRef]

22. Park, H.; Yamaguchi, R.; Imoto, S.; Miyano S. Uncovering Molecular Mechanisms of Drug Resistance via Network-Constrained Common Structure Identification. *J. Comput. Biol.* **2022**, *29*, 257–275. [CrossRef] [PubMed]

23. Wang, H.; Banerjee, A.; Boley, D. Common component analysis for multiple covariance matrices. In Proceedings of the Publication:KDD 11: 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 956–964.

24. Li, Y.; Luo, P.; Wu, C. A new network node similarity measure method and its applications. *arXiv* **2014**, arXiv:1403.4303.

25. Zou, H.; Hastie, T.; Tibshirani, R. Sparse principal component analysis. *J. Comput. Graph. Stat.* **2006**, *15*, 265–286. [CrossRef]

26. Li, C.; Li, H. Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics* **2008**, *24*, 1175–1182. [CrossRef]

27. Sun, H.; Lin, W.; Feng, R.; Li, H. Network-regularized high dimensional cox regression for analysis of genomic data. *Stat. Sin.* **2014**, *24*, 1433–1459. [CrossRef]

28. Park, H.; Yamaguchi, R.; Imoto, S.; Miyano S. Xprediction: Explainable EGFR-TKIs response prediction based on drug sensitivity specific gene networks. *PLoS ONE* **2022**, *17*, e0261630. [CrossRef] [PubMed]

29. Mizushima, T.; Ikeda, M.; Kato, T.; Ikeda, A.; Nishimura, J.; Hata, T.; Matsuda, C.; Satoh, T.; Mori, M.; Doki, M. Postoperative XELOX therapy for patients with curatively resected high-risk stage II and stage III rectal cancer without preoperative chemoradiation: A prospective, multicenter, open-label, single-arm phase II study. *BMC Cancer* **2019**, *19*, 929. [CrossRef]

30. Satake, H.; Yasui, H.; Kotake, T.; Okita, Y.; Hatachi, Y.; Kotaka, Y.; Kato, T.; Tsuji, A. First-line chemotherapy with capecitabine/oxaliplatin for advanced gastric cancer: A phase I study. *Mol. Clin. Oncol.* **2017**, *7*, 347–350. [CrossRef] [PubMed]

31. Quek, R.; Lim, W.; Foo, K.; Koo, W.; A-Manaf, A.; Toh, H. Capecitabine and oxaliplatin (XELOX) is safe and effective in patients with advanced gastric cancer. *Acta. Oncol.* **2007**, *46*, 1032–1034. [CrossRef]

32. Qu, Y.; Gao, N.; Wu, T. Expression and clinical significance of SYNE1 and MAGI2 gene promoter methylation in gastric cancer. *Medicine* **2021**, *100*, e23788. [CrossRef] [PubMed]

33. Ogony, J.; Choi, H.; Lui, A.; Cristofanilli, M.; Lewis-Wambi, J. Interferon-induced transmembrane protein 1 (IFITM1) overexpression enhances the aggressive phenotype of SUM149 inflammatory breast cancer cells in a signal transducer and activator of transcription 2 (STAT2)-dependent manner. *Breast Cancer Res.* **2016**, *18*, 25. [CrossRef] [PubMed]

34. Ahn, J.; Han, B.; Lee, M. Induction of Resistance to BRAF Inhibitor Is Associated with the Inability of Spry2 to Inhibit BRAF-V600E Activity in BRAF Mutant Cells. *Biomol. Ther.* **2015**, *23*, 320–326. [CrossRef]

35. Li, Y.; Umbach, D.; Krahn, J.; Shats, I.; Li, X.; Li, L. Predicting tumor response to drugs based on gene-expression biomarkers of sensitivity learned from cancer cell lines. *BMC Genom.* **2021**, *22*, 272. [CrossRef] [PubMed]

36. Sun, L.; Ke, M.; Wang, X.; Yin, M.; Wei, J.; Xu, L.; Tian, X.; Wang, F.; Zhang, H.; Fu, S.; et al. FAP high α-SMA low cancer-associated fibroblast-derived SLPI protein encapsulated in extracellular vesicles promotes ovarian cancer development via activation of PI3K/AKT and downstream signaling pathways. *Mol. Carcinog.* 2022, *online ahead of print*.

37. Sasaki, T.; Fujiwara-Tani, R.; Kishi, S.; Mori, S.; Luo, Y.; Ohmori H.; Kawahara, I.; Goto, K.; Nishiguchi, Y.; Mori, T.; et al. Targeting claudin-4 enhances chemosensitivity of pancreatic ductal carcinomas. *Cancer Med.* **2019**, *15*, 6700–6708. [CrossRef] [PubMed]

38. Yoshida, H.; Sumi, T,; Zhi, X.; Yasui, T.; Honda, K.; Ishiko, O. Claudin-4: A potential therapeutic target in chemotherapy-resistant ovarian cancer. *Anticancer Res.* **2011**, *31*, 1271–1277.

39. Breed, C.; Hicks, D.; Webb, P.; Galimanis, C.; Bitler, B.; Behbakht, K.; Baumgartner, H. Ovarian Tumor Cell Expression of Claudin-4 Reduces Apoptotic Response to Paclitaxel. *Mol. Cancer. Res.* **2019**, *17*, 741–750. [CrossRef]

40. Hicks, D.; Galimanis, C.; Webb, P.; Spillman, M.; Behbakht, K.; Neville, M.; Baumgartner, H. Claudin-4 activity in ovarian tumor cell apoptosis resistance and migration. *BMC Cancer* **2016**, *16*, 788. [CrossRef]

41. Nishiguchi, Y.; Fujiwara-Tani, R,; Sasaki, T.; Luo, Y.; Ohmori, H.; Kishi, S.; Mori, S.; Goto, K.; Yasui, W.; Sho, M.; et al. Targeting claudin-4 enhances CDDP-chemosensitivity in gastric cancer. *Oncotarget* **2019**, *10*, 2189–2202. [CrossRef]

42. Chen, Y.; Jiang, K.; Bai, X.; Liu, M.; Lin, S.; Xu, T.; Wei, J.; Li, D.; Xiong, Y.; Xin, W.; et al. ZEB1 Induces Ddr1 Promoter Hypermethylation and Contributes to the Chronic Pain in Spinal Cord in Rats Following Oxaliplatin Treatment. *Neuroch. Res.* **2021**, *46*, 2181–2191. [CrossRef]

43. Tao, Y.; Wang, R.; Lai, Q.; Wu, Q.; Wang,Y.; Jiang, X.; Zeng, L.; Zhou, S.; Li, Z.; Yang, T.; et al. Targeting of DDR1 with antibody-drug conjugates has antitumor effects in a mouse model of colon carcinoma. *Mol. Oncol.* **2019**, *13*, 1855–1873. [CrossRef]

44. Hur, H.; Ham, I.; Lee, D.; Jin, H.; Aguilera, K.; Oh, H.; Han, S.; Kwon, J.; Kim, Y.; Ding, K.; et al. Discoidin domain receptor 1 activity drives an aggressive phenotype in gastric carcinoma. *BMC Cancer* **2017**, *17*, 87. [CrossRef] [PubMed]

45. Menor, M.; Zhu, Y.; Wang, Y; Zhang, J.; Jiang, B.; Deng, Y. Development of somatic mutation signatures for risk stratification and prognosis in lung and colorectal adenocarcinomas. *BMC Med. Genom.* **2019**, *12* (Suppl. 1), 24. [CrossRef] [PubMed]

46. Chen, J.; Yuan, D.; Hao, Q.; Zhu, D.; Chen, Z. LncRNA PCGEM1 mediates oxaliplatin resistance in hepatocellular carcinoma via miR-129-5p/ETV1 axis in vitro. *Adv. Clin. Exp. Med.* **2021**, *30*, 831–838. [CrossRef] [PubMed]

47. Na, D.; Chae, J.; Cho, S.; Kang, W.; Lee, A.; Min, S.; Kang, J.; Kim, M.; Choi, J.; Lee, W.; et al. Predictive biomarkers for 5-fluorouracil and oxaliplatin-based chemotherapy in gastric cancers via profiling of patient-derived xenografts. *Nat. Commun.* **2021**, *12*, 4840. [CrossRef] [PubMed]

48. Giri, A. ETV5 expression positively correlates with promoter methylation and predicts response for 5-FU-based adjuvant therapy response in proximal colon cancer. *bioRxiv* **2020**. [CrossRef]

49. Borg, D.; Hedner, C.; Gaber, A.; Nodin, B.; Fristedt, R.; Jirstrom, K.; Eberhard, J.; Johnsson, A. Expression of IFITM1 as a prognostic biomarker in resected gastric and esophageal adenocarcinoma. *Biomark. Res.* **2016**, *58*, 10. [CrossRef]

50. Virag, P.; Fischer-Fodor, E.; Perde-Schrepler, M.; Brie, I.; Tatomir, C.; Balacescu, L.; Berindan-Neagoe, I.; Victor, B.; Balacescu, O. Oxaliplatin induces different cellular and molecular chemoresistance patterns in colorectal cancer cell lines of identical origins. *BMC Genom.* **2013**, *14*, 480. [CrossRef]

51. Lin, H.; Zhang, T.; Chen, M.; Shen, J. Novel biomarkers for the diagnosis and prognosis of gallbladder cancer. *J. Dig. Dis.* **2021**, *22*, 62–71. [CrossRef]

52. Mohanty, A.; Nam, A.; Pozhitkov, A.; Yang, L.; Srivastava, S.; Nathan, A.; Wu, X.; Mambetsariev, I.; Nelson, M.; Subbalakshmi, A.; et al. A Non-genetic Mechanism Involving the Integrin β4/Paxillin Axis Contributes to Chemoresistance in Lung Cancer. *iScience* **2022**, *23*, 101496. [CrossRef]

53. Penzvalto, Z.; Tegze, B.; Szasz, A.; Sztupinszki, Z.; Liko, I.; Szendroi, A.; Schafer, R.; Gyorffy, B. Identifying resistance mechanisms against five tyrosine kinase inhibitors targeting the ERBB/RAS pathway in 45 cancer cell lines. *PLoS ONE* **2013**, *8*, e59503. [CrossRef]

54. Riedesser, J.; Ebert, M. Precision medicine for metastatic colorectal cancer in clinical practice. *Ther. Adv. Med. Oncol.* **2022**, *14*, 17588359211072703. [CrossRef] [PubMed]

55. Zhang, Y.; Xia, F.; Liu, X.; Yu, X.; Xie, L.; Liu, L.; Chen, C.; Jiang, H.; Hao, X.; He, X.; et al. JAM3 maintains leukemia-initiating cell self-renewal through LRP5/AKT/β-catenin/CCND1 signaling. *J. Clin. Investig.* **2018**, *128*, 1737–1751. [CrossRef] [PubMed]

56. Zhang, Y.; Xu, Z.; Sun, Y; Chi, P.; Lu, X. Knockdown of KLK11 reverses oxaliplatin resistance by inhibiting proliferation and activating apoptosis via suppressing the PI3K/AKT signal pathway in colorectal cancer cell. *Onco. Targets Ther.* **2018**, *11*, 809–821. [CrossRef] [PubMed]

57. Hua, Q.; Li, T.; Liu, Y.; Shen, X; Zhu, X.; Xu, P. Upregulation of KLK8 Predicts Poor Prognosis in Pancreatic Cancer. *Front. Oncol.* **2021**, *11*, 624837. [CrossRef] [PubMed]

58. Zhang, Z.; Lee, J.; Lin, L.; Olivas,V.; Au, V.; LaFramboise, T.; Abdel-Rahman, M.; Wang, X.; Levine, A.; Rho, J.; et al. Activation of the AXL kinase causes resistance to *EGFR*-targeted therapy in lung cancer. *Nat. Genet.* **2012**, *44*, 852–860. [CrossRef]

59. Masica, D.; Karchin, R. Collections of simultaneously altered genes as biomarkers of cancer cell drug response. *Cancer Res.* **2013**, *73* 1699–1708. [CrossRef]

60. Luo, L.; McGarvey, P.; Madhavan, S.; Kumar, R.; Gusev, Y.; Upadhyay, G. Distinct lymphocyte antigens 6 (Ly6) family members Ly6D, Ly6E, Ly6K and Ly6H drive tumorigenesis and clinical outcome. *Oncotarget* **2016**, *7*, 11165–11193. [CrossRef]

61. AlHossiny, M.; Luo, L.; Frazier, W.; Steiner, N.; Gusev, Y.; Kallakury, B.; Glasgow, E.; Creswell, K.; Madhavan, S.; Kumar, R.; Upadhyay, G. Ly6E/K Signaling to TGFβ Promotes Breast Cancer Progression, Immune Escape, and Drug Resistance. *Cancer Res.* **2016**, *76*, 3376–3386. [CrossRef]

62. Burg, S. Correlates of immune and clinical activity of novel cancer vaccines. *Semin. Immunol.* **2018**, *39*, 119–136. [CrossRef]

63. Hu, T.; Zhang, Y.; Yang, T.; He, Q.; Zhao, M. LYPD3, a New Biomarker and Therapeutic Target for Acute Myelogenous Leukemia. *Front. Genet.* **2022**, *13*, 795820. [CrossRef]

64. Das, D.; Satapathy, S.; Siddharth, S.; Nayak, A.; Kundu, C. NECTIN-4 increased the 5-FU resistance in colon cancer cells by inducing the PI3K-AKT cascade. *Cancer Chemother. Pharmacol.* **2015**, *76*, 471–479. [CrossRef] [PubMed]

65. Jin, S.; Sun, Y; Liang, X.; Gu, X.; Ning, J.; Xu, Y.; Chen, S.; Pan, S. Emerging new therapeutic antibody derivatives for cancer treatment. *Signal Transduct. Target Ther.* **2022**, *7*, 39. [CrossRef] [PubMed]

66. Lu, W.; Fu, D.; Kong, X.; Huang, Z.; Hwang, M.; Zhu, Y.; Chen, L.; Jiang, K.; Li, X.; Wu, Y.; et al. FOLFOX treatment response prediction in metastatic or recurrent colorectal cancer patients via machine learning algorithms. *Cancer Med.* **2020**, *9*, 419–1429. [CrossRef]

67. Zhou, M.; Dong, J.; Huang, J.; Ye, W.; Zheng, Z.; Huang, K.; Pan, Y.; Cen, J.; Liang, Y.; Shu, G.; et al. Chitosan-Gelatin-EGCG Nanoparticle-Mediated LncRNA TMEM44-AS1 Silencing to Activate the P53 Signaling Pathway for the Synergistic Reversal of 5-FU Resistance in Gastric Cancer. *Adv. Sci.* **2022**, *9*, e2105077. [CrossRef]

68. Allert, C.; Waclawiczek, A.; Zimmermann, S.; Gollner, S.; Heid, D.; Janssen, M.; Renders, S.; Rohde, C.: Bauer, M.; Bruckmann, M.; et al. Maximilian Felix Blank 9 Protein tyrosine kinase 2b inhibition reverts niche-associated resistance to tyrosine kinase inhibitors in AML. *Leukemia* 2022, *online ahead of print*.

69. Zhang, Q.; Zhu, M.; Cheng, W.; Xing, R.; Li, W.; Zhao, W.; Xu, L.; Li, E.; Luo, G.; Lu, Y. Downregulation of 425G>a variant of calcium-binding protein S100A14 associated with poor differentiation and prognosis in gastric cancer. *J. Cancer Res. Clin. Oncol.* **2015**, *141*, 691–703. [CrossRef] [PubMed]

70. Feng, Y.; Wu, C.; Shiau, A.; Lee, J.; Chang, J.; Lu, P.; Tung, C.; Feng, L.; Huang, W.; Tsao, C. MicroRNA-21-mediated regulation of Sprouty2 protein expression enhances the cytotoxic effect of 5-fluorouracil and metformin in colon cancer cells. *Int. J. Mol. Med.* **2012**, *39*, 920–926.

71. Luo, J.; Chen, J.; Zhou, J.; Han, K.; Li, S.; Duan, J.; Cao, C.; Lin, J.; Xie, X.; Wang F. TBX20 inhibits colorectal cancer tumorigenesis by impairing NHEJ-mediated DNA repair. *Cancer Sci.* **2022**, *113*, 2008–2021. [CrossRef]

72. Tasaka, R.; Fukuda, T.; Shimomura, M.; Inoue, Y.; Wada, T.; Kawanishi, M.; Yasui, T.; Sumi, T. TBX2 expression is associated with platinum-sensitivity of ovarian serous carcinoma. *Oncol. Lett.* **2018**, *14*, 3085–3090. [CrossRef]

73. Esposito, A.; Bardelli, A.; Criscitiello, C.; Colombo, N.; Gelao, L.; Fumagalli, L.; Minchella, I.; Locatelli, M.; Goldhirsch, A.; Curigliano, G. Monitoring tumor-derived cell-free DNA in patients with solid tumors: Clinical perspectives and research opportunities. *Cancer Treat. Rev.* **2014**, *40*, 648–655. [CrossRef]

74. Sreekumar, R.; Al-Saihati, H.; Emaduddin, M.; Moutasim, K.; Mellone, M.; Patel, A.; Kilic, S.; Cetin, M.; Erdemir, S.; Navio, S.; et al. The ZEB2-dependent EMT transcriptional programme drives therapy resistance by activating nucleotide excision repair genes ERCC1 and ERCC4 in colorectal cancer. *Mol. Oncol.* **2021**, *15*, 2065–2083. [CrossRef]

75. Guo, Q.; Jing, F.; Xu, W.; Li, X.; Li, X.; Sun, J.; Xing, X.; Zhou, C.; Jing, F. Ubenimex induces autophagy inhibition and EMT suppression to overcome cisplatin resistance in GC cells by perturbing the CD13/EMP3/PI3K/AKT/NF-$\kappa\beta$ axis. *Aging* **2019**, *12*, 80–105. [CrossRef] [PubMed]

76. Zhou, X.; Men, X.; Zhao, R.; Han, J.; Fan, Z.; Wang, Y.; Lv, Y.; Zuo, J.; Zhao, L.; Sang, M.; et al. miR-200c inhibits TGF-β-induced-EMT to restore trastuzumab sensitivity by targeting ZEB1 and ZEB2 in gastric cancer. *Cancer Gene Ther.* **2018**, *25*, 68–76. [CrossRef] [PubMed]