Prediction method for therapeutic response at multiple time points of gene expression profiles

by Arika Fukushima

Graduate School of Life and Medical Sciences

Doshisha University

Contents

List of figures

List of tables

1. Introduction

Bioinformatics is a multidisciplinary field that spans medical, biological, pharmacological, and computational sciences [1]. Deciphering biological functions and exploring biomarkers are common research topics in this field [2][3]. Previously, only a single molecule or a select few molecules, such as tumor markers, have been used for disease diagnosis [4]. However, the advent of comprehensive molecular profiling techniques, such as those using the genome, transcriptome, proteome, and metabolome, has resulted in a paradigm shift in novel biomarker discovery. Conventional statistical approaches have limitations and suffer from the discovery of biomarker candidates in high-dimensional data [5]. Bioinformatics is a key technology for biomarker discovery in complicated and large-scale data.

Biomarkers that exhibit practical utility are classified as pharmacodynamic, prognostic, or predictive markers [6]. Pharmacodynamic biomarkers are often used for the rational and efficient development of new molecular therapeutics [7], such as Histone γH2AX and Poly (ADP-Ribose) [8]. Prognostic biomarkers provide the outcome of patients who do not depend on the treatment received, such as recurrence, disease progression, and death [9]. As a recent example of a prognostic biomarker, the transcriptome was Oncotype DX™ (Genetic Health). Oncotype DX™ provides patients with breast cancer with a recurrence score calculated from the expression profiles of 21 genes [10]. The predictive (or response) biomarkers evaluate whether patients benefit from a particular treatment [6]. One of the most well-known predictive biomarkers is PD-L1 for cancer [11]. In patients with triple-negative breast cancer, measuring the gene expression level of PD-L1 makes it possible to predict which are sensitive or not sensitive

to immune checkpoint inhibitors [12]. As described, prognostic and predictive biomarkers obtained by measuring gene expression profiles have helped determine treatment strategies.

The rapid development of transcriptome analysis in recent years has occurred owing to the evolution of measurement technology. One such technology is the deoxyribonucleic acid (DNA) microarray [13]. Microarrays use hybridization between fluorescent-labeled probes on glass substrates and samples to detect the fluorescent signals of ribonucleic acid (RNA) [14][15]. Because the microarray has thousands to millions of probes, it can comprehensively measure the expression profiles of large-scale genes in one experiment. However, microarrays have several limitations, including the high cost of measurement [16] and the small number of possible samples. Therefore, the main problem with microarray biomarker discovery was the discovery of some genes related to disease from a huge number of genes collected from small samples.

Large-scale genes and small samples make it challenging to discover biomarkers from microarrays [17][18]. This is referred to as the $p \gg n$ problem (p is the number of genes in microarray; n is the number of samples) [19]. There are statistical tests and machine learning as the methods for biomarker discovery, and, on both of those methods, the challenge for the $p \gg n$ problem persists.

Statistical tests are widely used as biomarker discovery methods for selecting genes with different gene expression profiles between two groups, such as case and control [2][20][21]. The t-test is one of the most popular and typical statistical tests. For the t-test, a p-value that indicates the probability that the null hypothesis that the two groups do not have different gene expression profiles is true was used. For biomarker discovery,

millions of pairs were created from large numbers of genes to compare gene expression profiles, and t-tests were performed for all pairs. However, repeating t-tests for these, all pairs sometimes occurred as false positives, called type I errors [22]. To reduce false positives, some methods to control the family-wise error rate [23] and false discovery rate [24], which are indicators of type I error, have been developed. Examples of these methods include Bonferroni [25], Holm [26], and Benjamini-Hochberg [27] methods. However, these methods do not entirely prevent false positives from occurring.

In the field of machine learning, the $p \gg n$ problem has an adverse effect called the curse of dimensionality [28][29]. A problem associated with dimensionality is that the number of samples required to construct a model without overfitting increases exponentially as the number of genes increases. However, increasing the number of samples is not easy because the cost of the microarray is too high to collect gene expression profiles from a large number of samples [30]. Therefore, to solve this problem using machine learning, a sparse modeling method such as the least absolute shrinkage and selection operator (lasso) approach has been utilized in many studies [31][32][33][34][35]. In sparse modeling methods, a model is constructed without over-fitting by using the hypothesis that a few genes are associated with the target biological phenomenon (in other words, genes that should be selected as a biomarker are sparse in microarray data). This hypothesis is suitable for biomarker discovery because some studies have suggested that some genes are significant biomarkers [36][37]. Thus, sparse modeling is helpful for biomarker discovery as a method that can select related genes from many genes in microarray data while avoiding the curse of dimensionality.

Biomarker discovery using microarrays is another challenge. Gene expression profiles often exhibit a strong correlation called multicollinearity [38][39]. This correlation might lead to the construction of an erroneous model using machine learning [40][41][42]. The variance inflation factor (VIF) is an indicator of multicollinearity. A gene pair with a VIF value of more than ten might have multicollinearity [43][44]. Before constructing the models, genes expressing multicollinearity of expression profiles should be removed from the microarray data. However, it is not clear which genes should be excluded from the gene pair, because VIF is insufficient for determining which genes should be excluded.

When there are gene expression profiles exhibiting multicollinearity, the lasso might fail to select genes. Lasso selects only one gene from a pair of genes with multicollinearity [45][46]. In sparse modeling, the elastic net is helpful for multicollinearity [47]. The elastic net is the method of the expanded the lasso to perform gene selection and simultaneously reduce the instability of multicollinearity. Therefore, the elastic net helps solve the $p \gg n$ problem (1) and address multicollinearity (2) for biomarker discovery.

Generally, microarray data for biomarker discovery are categorized into two types based on the collection method: a single time point and multiple time points (time-course) [48][49]. Because the gene expression profiles during the biological principles and processes are changed dynamically through multiple time points [50][51], time-course microarray data are more beneficial than those collected at a single time point [49][52]. Thus, the number of time-course gene expression profiles increased exponentially in the 2000s [53]. However, currently, the practical biomarker predicting the therapeutic response uses gene expression profiles at a single time point. For example, Oncotype

DX™ calculated the recurrence score of breast cancer using gene expression profiles at a single time point. Oncotype DX™ has the problem that the therapy decision for patients with a medium recurrence score is difficult [54][55], but this problem might be solved using time-course gene expression profiles. However, time-course gene expression profiles collected by microarray make biomarker discovery more complex and difficult than those of single-time points.

Biomarker discovery using time-course microarray data is problematized when expanding the search space of genes in proportion to the number of time points. Using microarray data of the single-time point, the number of gene expression levels for analysis is $p \times n$. However, their number in the time-course microarray data is $p \times n \times t$ (t is the number of time points), and the search space of time-course microarray data is t times larger than the single time point. Therefore, the time course microarray data make biomarker discovery more complicated than when using data from a single time point.

Another problem of biomarker discovery using time-course microarray data is that even in the same state of patients, the expression levels of genes as biomarkers continue to change dynamically *in vivo* [49]. For example, gene expression levels are different at each time point of observation by microarray, even though the patient state keeps to be diagnosed normal [56][57]. This problem prevents prognostic and predictive markers from spreading clinically because the dynamics of gene expression profiles reduce the accuracy of the markers. Therefore, prognostic and predictive markers need robustness to the dynamics of gene expression profiles.

Many studies of biomarker discovery that challenge dynamics often identify different genes at different time points as markers to clarify biological processing. For example,

the dynamical network biomarker utilized the changes in expressed genes between before and after disease state is changed [56][57]. This marker could contribute to clarifying biological processing. However, it is necessary to observe many genes as biomarkers before the disease state changes. Another example, F-Logistics was developed to identify genes with different dynamic alterations between normal and diseased states [50]. However, genes identified by F-Logistics sometimes have the same expression levels between normal and disease at a single time point because F-Logistics focuses on the shape of data formed by time-course gene expression levels. Thus, to maintain the accuracy of prognostic and predictive markers through multiple time points, gene expression levels of biomarkers need to be different between the two states at any time point.

To identify gene markers with high accuracy, we challenged the $p >> n$ problem (1), multicollinearity (2), and identification of genes with different expression levels between the two states at any time point (3). We expanded the elastic net to solve problems (1) and (2) using time-course microarray data. To solve the problem (3), we developed two approaches for identifying genes as prognostic and predictive markers.

First, we focused on genes exhibiting consistent gene expression levels between the two groups at all times. Studies have been conducted on time-course gene expression profiles [58][59][60][61], and most selected genes with time course patterns such as monotonous increase and decrease. However, these genes with different time-course patterns are sometimes unsuitable for biomarkers that are classified into two groups. There might be time points when gene expression levels are almost the same between the two groups at a certain point. Thus, we developed a new elastic net method called

"expanding elastic net for consistent differentiation" (eENCD) to select genes with consistent differentiation through time-course gene expression profiles to classify two groups at any time point.

The eENCD was designed to select genes with consistent differentiation at all time points (Figure 1). The genes whose expression levels were the same in the two groups at some points, such as Figure 1 b, might not be selected. This result means that genes with a large difference in expression levels in the two groups may be missed by eENCD. This limitation helps select candidate prognostic and predictive markers for a large number of genes. However, the genes selected by eENCD sometimes evaluate the biological process for the clinical application complex because genes strongly affected by therapy might be excluded. Therefore, the concept of biomarker discovery is also required to select genes with significant differences in expression levels between the two groups (4).

If the time-course gene expression profiles for prediction are used, many prediction results for as many time points will be derived. It is considered that this supports the decision of therapeutic strategy better than using a single time point because multiple predicted results provide more information about the target biological phenomenon. However, many prediction results make the decision difficult. Moreover, requiring many time points for prediction means that prediction in the early term becomes difficult. Therefore, the biomarkers must combine the results at multiple time points into one and predict the therapeutic response using a few time points (5).

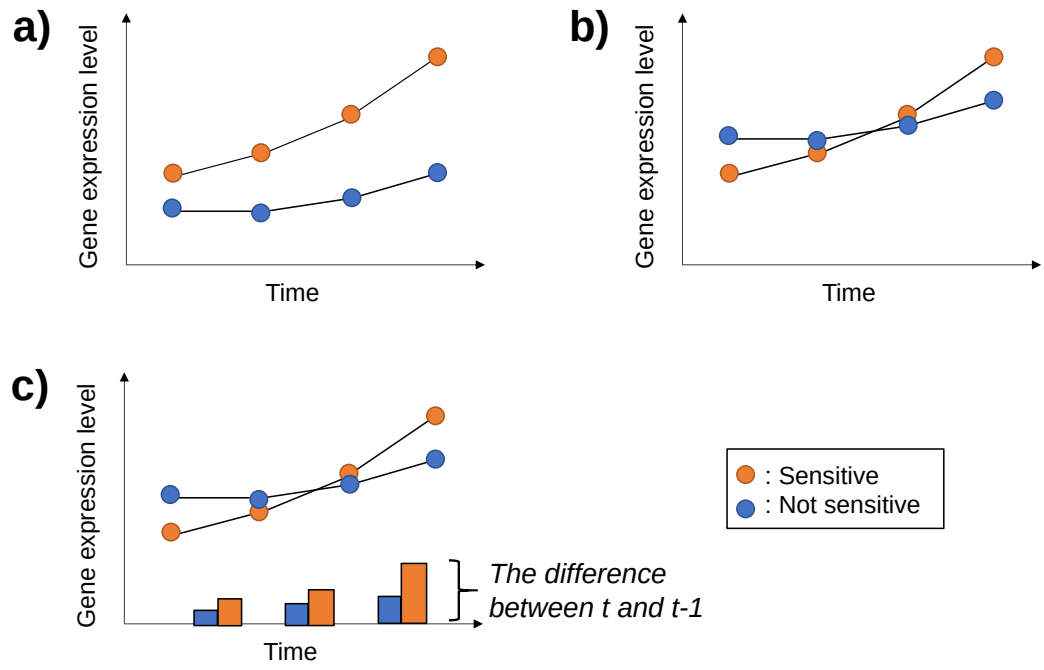

Figure 1 Expression levels of genes selected by each proposed method. (a) The eENCD selects genes showing consistent different gene expression levels between two groups, such as sensitive or not sensitive for therapy at all time points. (b) The eENCD does not select gene expression levels overlapping sensitive or not sensitive. (c) CPMTP selects genes with differently changed patterns of expression levels in the two groups at all time points. CPMTP can do gene expression levels that overlap between sensitive and not sensitive because CPMTP utilizes the difference between t and t-1 gene expression levels.

Second, a new elastic net method was proposed to solve problems (1)–(5). With this new method, we solved the problem (4) using the gene expression levels from one point to another. Using the difference between time points made it possible to select genes with

different patterns of expression levels because the difference became zero if the pattern of gene expression levels was changed, such as the same between the two groups. Moreover, to solve the problem (5), we utilized the Bayesian theorem to integrate predictive results into one, with the hypothesis that more time points were used and more accurate classification was performed. This is because other biomarkers that dynamically update predictive indicators with increasing time points accurately classify the two groups [62][63]. This new method, called "consolidating probabilities of multiple time points" (CPMTP), utilizes consolidating probabilities calculated with the difference of gene expression profiles and a model constructed using the elastic net at multiple time points.

This study introduced and evaluated two approaches for biomarker discovery: eENCD and CPMTP (Table 1). The second section described the eENCD by providing details on the background and algorithm, and the method was evaluated using two datasets of gene expression profiles collected from multiple sclerosis patients at multiple time points. The third section described the CPMTP. Because CPMTP includes a gene selection method (CPMTPg) and a unique prediction method (CPMTPp), the algorithms associated with each were described separately. CPMTP was evaluated using two time-course microarray datasets collected from patients with MS and hepatitis C virus (HCV). The fourth section discusses biomarkers in terms of accuracy and the genes selected by eENCD and CPMTPg. The final section presents the conclusion.

Table 1 Problems and solutions in eENCD and CPMTP

No.	Problem	eENCD	CPMTP
(1)	$p >> n$ problem	elastic net	elastic net
(2)	multicollinearity	elastic net	elastic net
(3)	selection of genes having different gene expression levels at multiple time points	utilization of the weights in elastic net for selecting the genes common to multiple time points.	utilization of the optimization method (Ex. Genetic algorithm) for tuning the parameters in the prediction methods for the problem (5)
(4)	selection of genes overlapping gene expression levels between sensitive and not sensitive	--	utilization of difference of gene expression levels between one and the next time points
(5)	integration of multiple results at each time point for prediction	--	consolidating predicted results using Bayesian theory

The eENCD challenged problems (1)–(3). However, eENCD did not solve problems (4) and (5). Therefore, we developed CPMTP to solve these problems.

2. Expanding elastic net for consistent differentiation: eENCD

2.1. Background of eENCD

MS is one of the most common neurological disabilities of the central nervous system [64]. The highest incidences of MS have been reported in North America and Europe (100/100,000), and the lowest in East Asia and sub-Saharan Africa (2/100,000) [65]. This disease is the second most common neurological disability in young adulthood [66]. Approximately 80–90% of MS patients initially suffer from relapsing-remitting MS (RRMS) in which MS repeatedly occurs with symptoms, including the stages of neurological disability (relapse) and recovery [64]. The disease gradually shifts to secondary progressive MS which is associated with frequent relapses. Therefore, a systematic treatment strategy to prevent and/or delay relapse is important for the improvement of the quality of life of MS patients.

Interferon-β (INF-β) has been commonly used to prevent the relapse of MS [67][68]. However, INF-β treatment has two issues. First, the treatment only works for a limited number of patients, where approximately half of the patients relapse within two years despite treatment [69][70]. Second, this treatment can cause side effects, such as spasticity and dermal reaction [68]. Thus, effective surveillance and appropriate intervention over a long period post-treatment are required. Although the pathogenesis of MS has not yet been fully elucidated, various genetic factors involved in this disease have been reported [71]. Gene expression data have been intensively analyzed to predict INF-β treatment

response [67][68][71][72][73][74][75]. Hundreds of genes, such as *Caspase2*, *Caspase10*, and *FLIP* genes, showed promise in predicting the therapeutic response [68][74]. However, these genes were identified by the conventional statistical method which showed low prediction accuracies in some cases [74][75]. The *MxA* and *ISG* genes were reported to be predictive for IFN-β therapeutic response [71][72][73]. However, these gene expression levels were not consistently differentiated over all of the time-course, which means that the prediction would be accurate immediately after the observation of the gene expression levels, but the accuracy of prediction would be low for subsequent responses [72]. Therefore, the identification of genes showing highly accurate prediction abilities throughout all time-courses is needed.

Generally, data analyses to identify biomarkers are categorized into the single-time point and time-course-based approaches. Prediction using only the currently observable data to predict an outcome of treatment is the most useful but challenging for optimizing patient treatment. However, single-time point-based analyses [48] (Figure 2 a) are a challenge since the gene expression levels during the progression of MS are dynamic [49][50][51]. Prediction using time-course data, that is multiple time point data, would result in more accurate predictions [50][58][59][61][60][76] by eliminating the selection of genes showing inconsistent differentiation over the observation period (Figure 2 b and c). There are several difficulties in microarray data analyses, including the problem of high-dimensionality (a higher number of genes compared to sample size). Sparse modeling methods have been commonly utilized to identify differentiated genes to

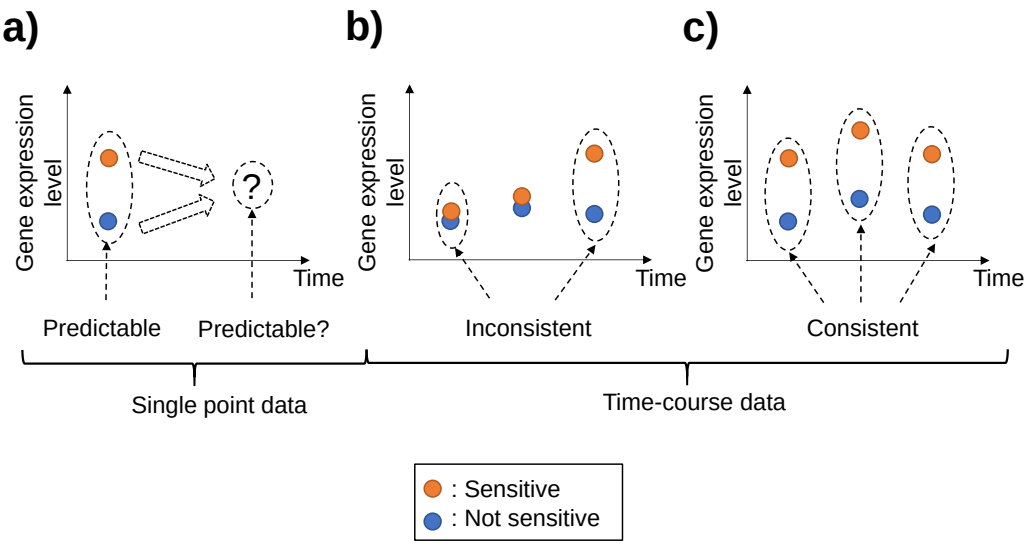

Figure 2 The concept of prediction using gene expression profiles. (a) Genes show the current expression levels at a single time point. (b) Genes show inconsistent differentiation between the current and the future time points. (c) Genes show consistent differentiation throughout the data at multiple time points.

address this issue [50][77][78]. However, to our knowledge, these methods were designed to analyze single-time point data analyses and their application for time-course data analyses has not been reported.

The purpose of this study was to develop a novel analytical method to efficiently identify genes showing consistent differentiation throughout the time-courses by modifying sparse modeling methods. As an application, two microarray time-course datasets collected from patients with MS were used for predictions of INF-β treatment responses, and the prediction accuracies of eENCD were evaluated using the conventional method.

2.2. Selecting method by eNCD

The elastic net [47], a sparse modeling method, was modified to analyze time-course data. Our method was designed to find genes showing consistent differentiation between the two given groups throughout multiple time points. Here, we addressed the following problems:

(1) High dimensionality. Microarray data include a larger number of genes compared to small sample size.
(2) Multi-collinearity. Microarray data include many genes showing highly positive correlations. The use of these genes for a prediction model would deteriorate its generalization ability [35].
(3) Time-courses. Genes showing consistent differentiation throughout multiple time points should be identified.

The elastic net was designed to analyze single-time point data for identifying differentiated genes by preventing multi-collinearity [47]. We modified this method for the time-course data analyses.

2.2.1. Background techniques of eENCD

Sparse modeling is one of a variety of selection methods suitable for high-dimensional data analyses [47][79]. Among the different sparse modeling methods, the lasso has been commonly used in various studies [50][77][78][80]. However, the lasso has a limitation; this method selects only one variable from two variables showing a high correlation (multicollinearity), where the other variables are not selected despite being differentiated [47]. The ridge regression model is a method capable of solving this problem. This method can construct models with two variables showing a multi-collinearity: however, this method does not select genes. The elastic net is another sparse modeling method able to reduce those two limitations. The elastic net is comprised of the lasso and ridge [35], which selects variable sets, that is this method selects all variables, even those showing high multi-collinearities [47][81][82]. Here, we employed the elastic net rather than the lasso to select gene candidates showing predictive abilities for subsequent analyses.

The eENCD used logistic regression (Eq. 1) to predict the INF-β treatment response based on differentiated genes [33][50][81].

$$\Pr(\mathbf{y} = 1 | \boldsymbol{X}_{\boldsymbol{t_r}}) = \frac{1}{1 + e^{-(\boldsymbol{X}_{\boldsymbol{t_r}} \boldsymbol{\beta}_{\boldsymbol{t_r}})}}$$

Eq. 1

where

$\mathbf{y} = \{y_1, y_2, \ldots, y_n;\ y_i \in \{0,1\}\}$: y_i denotes the response variable which included sensitive responder (labeled as 1) or not sensitive responder (labeled as 0) for INF-β treatment, respectively. *n* denotes the sample size of MS patients.

$\boldsymbol{X}_{t_r} = \{\mathbf{x}_{t_r,1}, \mathbf{x}_{t_r,2}, \ldots, \mathbf{x}_{t_r,p};\ \mathbf{x}_{t_r,j}{}^{\top} = \{x_{t_r,j}^{(1)}, x_{t_r,j}^{(2)}, \ldots, x_{t_r,j}^{(n)};\ j = 1, \ldots, p\}\}$: $\boldsymbol{X}_{t_r}$ denotes the explanatory variables of gene expression levels at time point t_r $(r = 0, \ldots, R)$. p denotes the number of genes.

$\boldsymbol{\beta_{tr}} = \{\beta_{t_r,1}, \beta_{t_r,2}, \ldots, \beta_{t_r,p}\}$: $\boldsymbol{\beta_{tr}}$ denotes the regression coefficients at time point t_r.

The regression coefficient $\boldsymbol{\beta_{tr}}$ in Eq. 1 indicates the degree of association between the response to INF-β treatment and each gene. Therefore, a gene with a high absolute value of a regression coefficient was selected as a gene bearing the predictive ability of the therapeutic response. However, regression coefficients were difficult to calculate by Ordinary Least Squares (OLS), which is a general method for calculation of the regression coefficients, because of the high dimensionality of the microarray data. Therefore, a small number of differentiated genes should be selected before the use of OLS. Sparse modeling had an assumption that only several regression coefficients were needed for the prediction model and that the others were not needed. This assumption meant that the regression coefficient values of several genes which were needed for the prediction model were non-zero while the other values were zero. In short, genes with regression coefficients of non-zero were selected as genes to predict the responses to INF-β treatment. With the use of the elastic net, regression coefficients were calculated by adding a penalty term to a least-square loss function (Eq. 2).

$$\underset{\boldsymbol{\beta_{tr}}}{\operatorname{argmin}} \mathrm{J}(\boldsymbol{y}, \boldsymbol{X_{tr}}) \; + \; \lambda \sum_{j=1}^{p} w_{t_r,j} \left[(1-\alpha) \frac{1}{2} \beta_{t_r,j}^2 + \alpha \left| \beta_{t_r,j} \right| \right]$$

Eq. 2

where

$\mathrm{J}(\boldsymbol{y}, \boldsymbol{X_{t_r}})$ denotes loss of function of OLS; λ denotes the hyper-parameter for the penalty term of the elastic net. The penalty term was given after the second term of the equation. Hyper-parameters were generally set by analysts.

α $(0 \leq \alpha \leq 1)$ denotes the hyper-parameter that indicated the degree between the ridge $(\frac{1}{2}\beta_{t_r,j}^2)$ and the lasso $(\left|\beta_{t_r,j}\right|)$ terms.

$\boldsymbol{w_{t_r}} = \{w_{t_r,1}, w_{t_r,2}, \ldots, w_{t_r,p}\}$ ($\boldsymbol{w_{t_r}} \in \mathbb{R}_{>0}$): $\boldsymbol{w_{t_r}}$ denote the weights of the elastic net as the selection bias of each gene at a time point t_r. Thus, a gene with a larger and a lower weight was selected in a lower and a higher probability, respectively.

Cross-validation is commonly used for optimizing the λ value in Eq. 2. However, inconsistent genes are usually selected depending on the λ value. To prevent this problem, the stability selection was used here instead [83]. The stability selection selects for genes according to the following procedures:

(i) A subset of samples was yielded from the gene expression data by random sampling.
(ii) An arbitrary λ value was given to the elastic net to select genes using the data of (i).
(iii) (i) ~ (ii) were repeated with multiple subsets.

(iv) The frequency of selection with an arbitrary λ value was calculated for multiple subsets.

(v) (i) ~ (iii) were repeated with multiple λ values.

(vi) For each gene, the maximum probability calculated in (iv) among multiple λ values was regarded as the selection probability of the gene.

(vii) Genes showing a selection probability (above the threshold θ_{ss}) were selected.

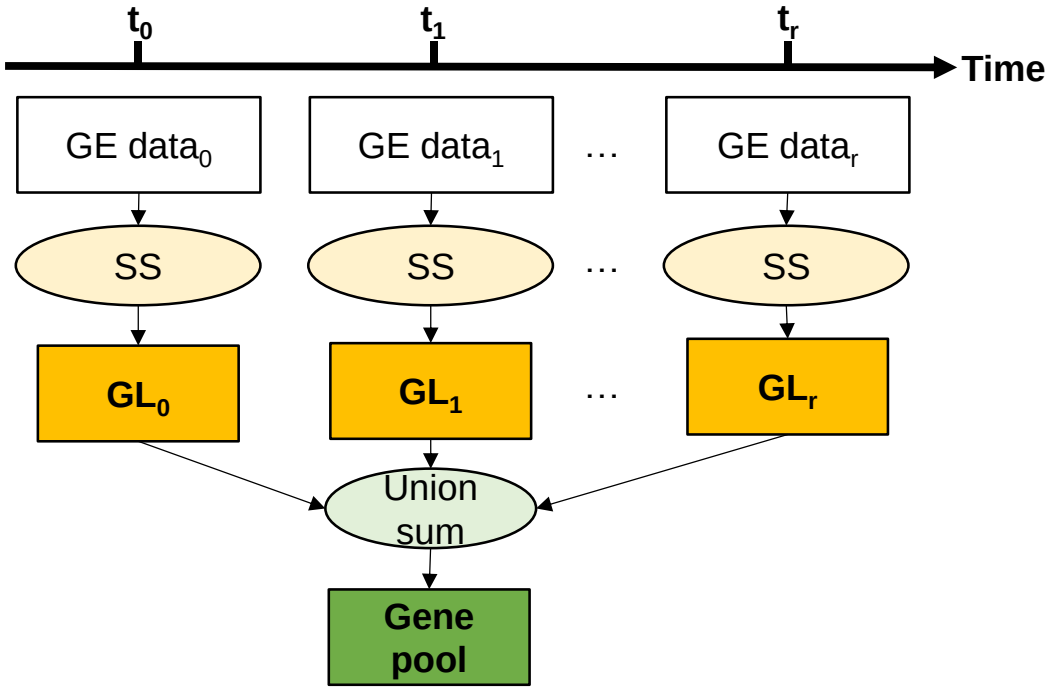

Figure 3 The concept of eENCD: Screening step. The gene pool is created by stability selection (SS) using gene expression data at each time point. $\boldsymbol{GL_r}$ is the gene list selected by SS at the time point t_r.

2.2.2. Algorithm of eENCD

The eENCD consisted of the following three procedures:

(i) Screening of gene candidates (Figure 3). Due to the difficulties associated with high dimensional problems, the elastic net with SS was used for the screening of gene candidates, known as the gene pool, from the data at each time point. Only genes selected at least one time by the elastic net with the stability selection were selected in the gene pool and the rest were eliminated.

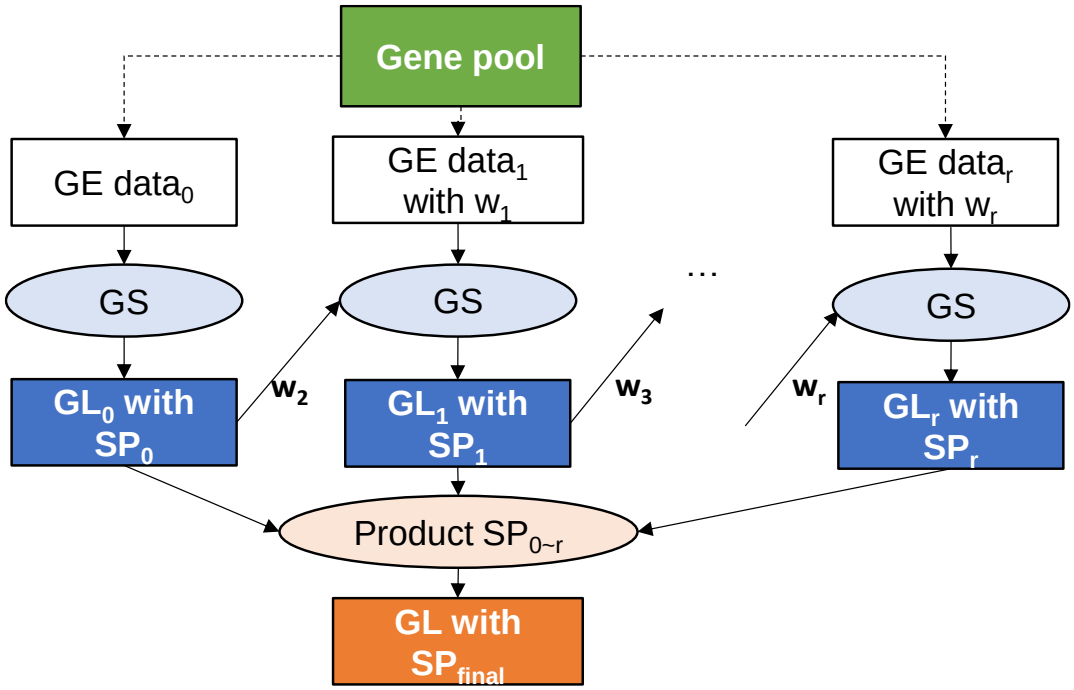

Figure 4 The concept of eENCD: Gene selection (GS) by candidate genes and calculating selected probability (SP). Gene selection (GS) uses elastic net with stability selection assigning weights ($\boldsymbol{w_1}$~$\boldsymbol{w_r}$) to gene expression data at each time point. SP is a selected probability.

(ii) Ranking genes showing consistent differentiation throughout multiple time points (Figure 4). Modified the elastic net was used to select genes showing consistent differentiation throughout multiple time points from the gene pool. Firstly, the elastic net with the stability selection selected predictive genes from the gene pool at the first time point. Secondly, at the next time point t_r, the elastic net (Eq. 2) with the stability selection was conducted with a higher selection bias to select genes that were selected at the previous time point $r-1$. Therefore, the elastic net sets the weights for which genes selected at $r-1$ were smaller than genes not selected (Eq. 3). This procedure was repeatedly performed at subsequent time points.

Consequently, genes showing consistent differentiation throughout multiple time points were identified:

$$w_{t_r,j} = \begin{cases} 1, & \text{if } g_j \in \boldsymbol{GL}_{t_{r-1}} \\ \gamma, & \text{if } g_j \notin \boldsymbol{GL}_{t_{r-1}} \end{cases}$$

Eq. 3

where

$w_{t_r,j}$: the weight of the j^{th} gene in the elastic net at a time point t_r in Eq. 2.

γ ($\gamma \in \mathbb{R}_{>0}$; $\gamma > 1$): the selection bias.

g_j denotes the j^{th} gene.

$\boldsymbol{GL}_{t_{r-1}}$: a gene list at t_{r-1}. The gene list was constructed with selected genes at t_{r-1}.

The product SP_{final} denoted the probabilities based on the frequency of selection of each gene throughout all time points (Eq. 4). The product SP_{final} was ranked in descending order. According to this ranking, the gene list for the prediction model was created to be used for the model in the third step:

$$\boldsymbol{SP_{final}} = \{SP_1, SP_2, \ldots, SP_p\}$$

$$SP_j = \prod_{r=1}^{R} SP_{t_r,j}$$

Eq. 4

Where

$\boldsymbol{SP_{final}}$: the product was calculated by selection probability at each time point for each gene, and the genes were ranked in descending order of the selection probabilities.

$SP_{t_r,j}$: selection probability of the j^{th} gene by the stability selection at a time point t_r.

(iii) Construction of a prediction model using the ranked genes (Figure 5). Genes for the prediction model were identified based on the gene list ranked in the second step. The time point of data to be used for constructing the prediction model was also selected simultaneously. Here, prediction models of therapeutic response were constructed with combinations of various groups of genes and time points of gene expression data. To identify the genes and select a time point of data for the prediction model, these models were evaluated. An evaluation value was calculated by the one prediction model which was constructed by one group of genes using time point data. The genes in the group with the best evaluation value were identified as the genes showing consistent differentiation. These time point data were selected for the prediction model. The group of genes was created by adding one by one from the gene list of the second step in descending order. The prediction models of all gene groups were constructed and evaluated at each time point of the time-course data.

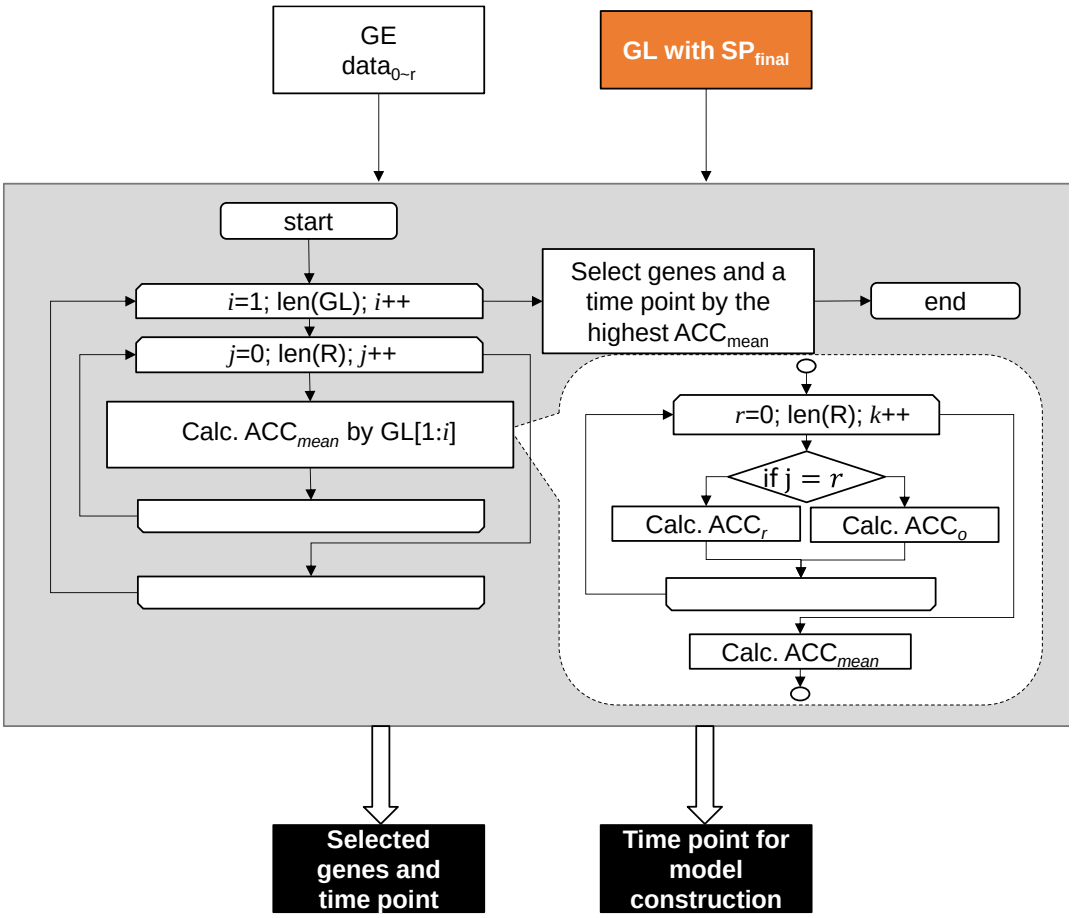

Figure 5 The concept of eENCD: Identifying the genes from gene lists with SP. In this step, prediction models are evaluated using leave-one-out cross-validation (time point for model construction = time point for prediction) and utilizing different data (time point for model construction ≠ time point for prediction).

In this step, prediction accuracy, a ratio that the prediction model accurately predicted using the data which were not used for constructing the model, was used as an evaluation value. A prediction model was constructed by a group of genes using data at a time point,

and the prediction accuracy was calculated. Prediction accuracies were calculated at each time point for model construction, as shown in the following two cases.

Case 1: Time point for model construction = time point for prediction

A constructed model was used for the prediction of the data at an identical time point. Leave-one-out was used to evaluate the prediction accuracy ACC_o. In the leave-one-out cross-validation, one sample of data was used as test data, and other data were used for model construction. The leave-one-out cross-validation was repeated until all the samples became test data.

Case 2: Time point for model construction ≠ time point for prediction

A constructed model was used for the prediction of the data at a time point not used for the construction. The prediction accuracies ACC_R were calculated using the data at time points for prediction.

The mean of prediction accuracies ACC_{mean} were calculated for each group of genes and for each time point of model construction in Eq. 5.

$$\mathrm{ACC}_{mean} = \frac{1}{R}\left(\sum_{d \in \boldsymbol{D}} ACC_R^{(d)} + ACC_o\right)$$

Eq. 5

where

ACC_{mean}: the mean of the prediction accuracy for the prediction model constructed by a group of genes and a time point of data.

R: the length of all time points.

$ACC_R^{(d)}$: the prediction accuracy using data at time point d.

$\boldsymbol{D} = \{t_0, \dots, t_d, \dots, t_R\}$: the time points of $ACC_R^{(d)}$. t_d was not included in the time point used for model construction.

ACC_o: the prediction accuracy using data at the time point used for model construction.

The genes in the group with the best ACC_{mean} were identified as the selected genes. This model was constructed using the data of a time point. This time point was selected as the time point for model construction.

2.2.3. Numerical experiments for eENCD

The prediction accuracies of the developed models were evaluated for the prediction of INF-β treatment responses. The prediction accuracies of eENCD and the conventional methods were compared.

Material and pre-processing:

The evaluated data consisted of the time-course gene expression data in two MS patients who underwent INF-β treatment. The two datasets of GSE19285 (dataset A) [84] and GSE24427 (dataset B) [85] were used. Table 2 shows the number of time-course points in each data platform, and the method of normalization. Log_2-fold change and quantile normalization were performed as the pre-processing of gene expression data. Subsequently, the expression levels of each gene were converted to *Z*-scores. In this paper, a good response was treated as sensitive, and a poor response was treated as not sensitive.

Conventional method:

The conventional method used only the gene expression data at a single time point. elastic net with the stability selection using data at a single time point was used as the conventional method. Genes were ranked according to the selection probabilities by the stability selection. Finally, using the procedures of eENCD, ACC_{mean} was calculated using these selection probabilities. Thereafter, the genes in the group with the best ACC_{mean} were regarded as identified genes. These genes were regarded as genes with

Table 2 Summary of gene expression datasets of INF-β treatments for MS patients.

Name of dataset	Dataset A	Dataset B
GEO ID	GSE19285	GSE24427
Type of INF-β	Intramuscular INF-β1a	Subcutaneous INF-β1a
Time points	first (t_0), Second (t_1), fifth (t_2)	first (t_0), Second (t_1), 1 month (t_2), 12 month (t_4), 24 month (t_4)
Number of sensitive responders	15	16
Number of not sensitive responders	9	9
Number of genes	11220	13513
Gene expression	Peripheral blood mononuclear cells	Peripheral blood mononuclear cells
Platform	Affymetrix Human Genome U133A Array	Affymetrix Human Genome U133A Array
Preprocessing for microarray	MAS5.0	MAS5.0

The symbols of time points are presented t_0, t_1, t_2, etc.

the best performance throughout multiple time points in the conventional method using data at a single time point.

Evaluation method:

The prediction accuracies were calculated by Eq. 6. These were calculated using only test data which were not used for model construction. To evaluate the prediction accuracies using the data at the time point used for model construction, leave-one-out cross-validation was conducted. To evaluate the prediction accuracies using the data at the other time points, all available data were used.

$$\mathrm{ACC}\ [\%] = \frac{TP + TN}{TP + FP + FN + TN} * 100$$

Eq. 6

where

TP: the number of true positives in the test data.

FP: the number of false positives in the test data.

FN: the number of false negatives in the test data.

TN: the number of true negatives in the test data.

First, the prediction accuracies of the construction models were evaluated. To compare the prediction model of eENCD and the conventional method, the mean prediction accuracy (ACC_{mean}) throughout all time points was calculated using ACC_R and ACC_o at each time point using Eq. 5. The lowest prediction accuracy, that is the minimum prediction accuracy (ACC_{min}) throughout all time points, was selected from ACC_R and ACC_o at each time point. To access the specificity and the sensitivity at each time point,

the receiver operating characteristic (ROC) curves, the area under the ROC curve (AUC), and the 95% confidence intervals were calculated.

Secondly, bootstrap sampling was performed to evaluate the prediction accuracies of selected genes at time points that had not been selected for model construction [86]. Bootstrap sampling selected samples from all samples by random sampling with replacement. A prediction model was constructed using selected samples and selected genes by either eENCD or conventional method. Prediction accuracies were calculated by each data at different time points; this was then repeated. Finally, the mean and standard deviation were calculated using the prediction accuracies for each subset at different time points. The difference in the prediction accuracies between the conventional method and eENCD was tested using the Student's *t*-test.

Thirdly, the differences between sensitive and not sensitive responders in the expression levels of the genes of eENCD were investigated. The expression levels of these genes at each time point were classified into two groups according to the therapeutic response, and the differences between the two groups were tested using the Wilcoxon rank-sum test. The p values in the Wilcoxon rank-sum test were adjusted by the Benjamini-Hochberg method (BH method). Furthermore, the median values of the expression levels of genes at each time point were compared between sensitive and not sensitive responders to assess whether the expression levels of the two groups were consistently different throughout all time points. The names of the selected genes were obtained from Gene Cards (http://www.genecards.org/).

Parameter and implementation:

The number of iterations for the stability selection was 500. Random sampling in the stability selection included 12 sensitive and 8 not sensitive responder samples, which were common to datasets A and B. The λ values, a hyper-parameter of the stability selection, were changed from 0.01 to 1.00 in 0.01 increments. The threshold θ_{ss} was 0.5. The hyper-parameter of the elastic net in Eq. 2 was $\alpha = 0.5$, and the weight parameter in Eq. 3 was $\gamma = 2$. The parameters of the elastic net with the stability selection used in the conventional method were also the same as the parameters in eENCD. The responses at 24 months after INF-β treatment in datasets A and B (Table 2) were predicted. Bootstrap sampling in the evaluation was repeated 50 times per prediction model at a different time point, and the prediction accuracies were calculated.

For the implementation of numerical experiments, R language (ver. 3.2.5; https://cran.r-project.org/bin/windows/base/old/3.2.5/) was used, wherein the "*limma*" and "*glmnet* (ver. 2.0-5)" packages were used for quantile normalization and the elastic net, respectively.

2.3. Results of eENCD

The eENCD and conventional methods were evaluated by the analyses of datasets A and B. The genes showing the best ACC_{mean} were identified for each dataset using both eENCD and the conventional method. The prediction accuracies at each time point and their mean in the first evaluation are listed in Table 3 and Table 4.

As an analytical result of dataset A, eENCD identified 11 genes and constructed the prediction model using the t_0 data. With the conventional method, prediction models were constructed using the t_0, t_1, and t_2 data, from which 9, 8, and 21 genes were identified, respectively.

Table 3 showed the prediction accuracies at each time point and their mean. The ACC_{mean} and ACC_{min} values using eENCD were 86% and 79%, respectively. With the conventional method using t_1 data, the ACC_{mean} was 86% and was comparable to that with eENCD. However, the ACC_{min} by the conventional method using the t_1 data was only 67% and was lower than that by eENCD. The ACC_{mean} by the conventional method using the t_0 and t_2 data was 83% and 79%, respectively. The ACC_{mean} by eENCD was higher than those of the conventional method. Here, we focus on the results at different time points in the first evaluation. The prediction accuracies by the proposed method were 92% at t_1 and 79% at t_2. The conventional method using t_1 data could predict therapeutic responses at t_2 with 92% accuracy; however, all the other results were lower than those of eENCD.

Table 3 Accuracy of prediction models by eENCD and conventional methods with dataset A.

Method	Accuracy [%]			
	t_0	t_1	t_2	Mean (ACC_{mean})
eENCD	(88)	***92***	79	***86***
Conventional method	(100)	71	79	83
	67	(100)	***92***	***86***
	54	83	(100)	79

Table 4 Accuracy of prediction models by eENCD and conventional methods with dataset B.

Method	Accuracy [%]					
	t_0	t_1	t_2	t_3	t_4	Mean (ACC_{mean})
eENCD	(96)	72	**92**	**84**	**76**	**84**
Conventional method	(92)	68	84	76	64	77
	72	(84)	76	60	64	71
	72	**92**	(96)	80	64	81
	72	64	68	(100)	40	69
	68	60	76	72	(96)	74

The number in parentheses is calculated by leave-one-out at the time point of data used by the prediction model. ***Bold accuracy*** is a top accuracy of each time point, but the top accuracy of t_0 is not presented since gene expression data at t_0 is used data by the proposed method. An accuracy of under line is the minimum accuracy (ACC_{min}) of each method.

As a result of the use of dataset B, eENCD identified 8 genes and constructed the prediction model using the t_0 data. The conventional method identified 5, 19, 7, 6, and

19 genes using t_0, t_1, t_2, t_3, and t_4 data, respectively.

Table 4 lists the prediction accuracies at each time point and the ACC_{mean}. The ACC_{mean} and ACC_{min} of eENCD were 84% and 72%, respectively. The ACC_{mean} values were 77%, 71%, 81%, 69%, and 74% using t_0, t_1, t_2, t_3, and t_4 data for model construction by the conventional method, respectively. The ACC_{min} values were 64%, 60%, 64%, 40%, and 60% using t_0, t_1, t_2, t_3, and t_4 data for model construction by the conventional method, respectively. The ACC_{mean} and ACC_{min} values of eENCD were higher than those of the conventional method. We focus on the results at different time points in the first evaluation. The prediction accuracies of eENCD using t_0 data were 92%, 84%, and 76% at time points t_2, t_3, and t_4, respectively. The prediction accuracy at t_1 by the conventional method using t_2 data was 92% and higher than that by eENCD. The other accuracies by eENCD were higher than those of the conventional method except for one case.

Bootstrap sampling was performed to evaluate the prediction accuracies at different time points in the second evaluation. Figure 6 shows the mean and standard deviation of prediction accuracies by eENCD and conventional methods at different time points. As shown in Figure 6 a, in dataset A, the mean accuracy of the different time points (t_1 and t_2) was 81%. This prediction accuracy was significantly higher than 65% ($\mathrm{p} = 2.06 \times 10^{-23}$), 71% ($\mathrm{p} = 1.48 \times 10^{-10}$), and 68% ($\mathrm{p} = 1.16 \times 10^{-16}$) at t_0, t_1, and t_2 in the conventional method ($p < 0.001$), respectively. As shown in Figure 6 b, in dataset B, the mean accuracy of the different time points (t_1, t_2, t_3, and t_4 by eENCD was

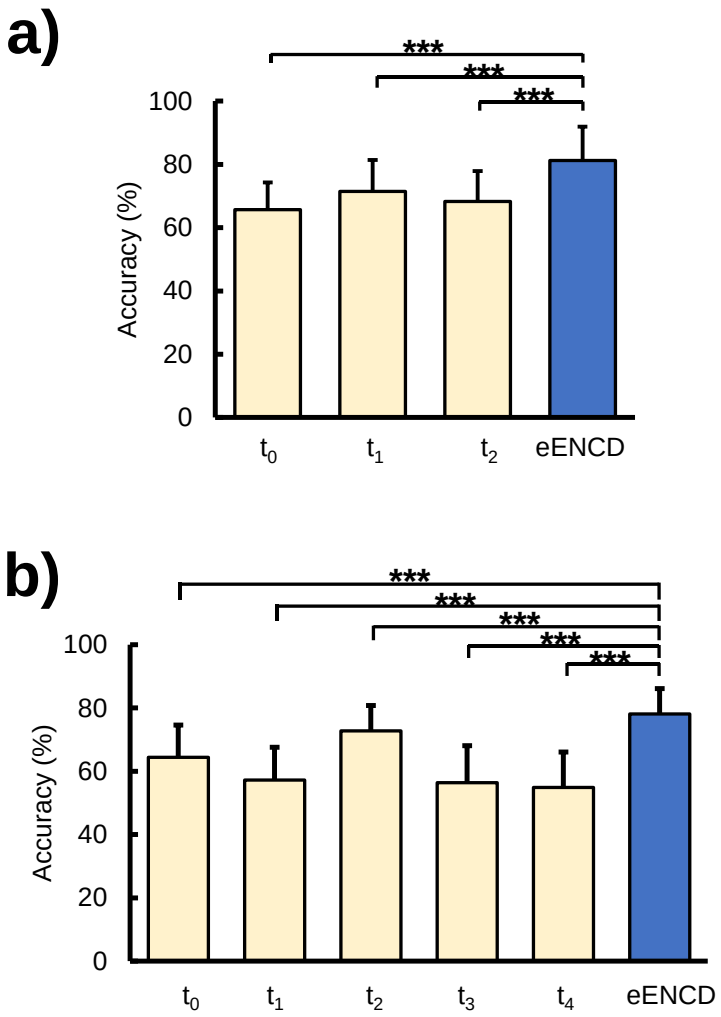

Figure 6 Prediction accuracies of eENCD by bootstrap sampling. The "***" means that the p-value of the student's t-test is less than 0.001. (a) Prediction accuracies for dataset A. The accuracies are the mean accuracies of different time points obtained without using the prediction model. (b) Prediction accuracies for dataset B. As with (a), the accuracies are the mean accuracies.

78%. The prediction accuracy by the conventional method at t_0, t_1, t_2, t_3, and t_4 were 64% ($p = 2.41 \times 10^{-40}$), 57% ($p = 1.73 \times 10^{-94}$), 73% ($p = 8.70 \times 10^{-11}$), 56%

($p = 1.30 \times 10^{-103}$), and 55% ($p = 1.46 \times 10^{-78}$), respectively. In dataset B, the mean accuracy of the different time points by eENCD was significantly higher than those by the conventional method ($p < 0.001$). Therefore, the prediction accuracies at different time points by eENCD were significantly higher than those by the conventional method.

To assess the sensitivity and specificity of the prediction model by eENCD, ROC curves and AUCs in datasets A and B were measured, as shown in Figure 7 a and b. As shown in Figure 7 a, in dataset A, the AUCs at t_0, t_1, and t_2 were 0.95, 0.94, and 0.90 by eENCD, respectively, all of which were higher than or equal to 0.9. The lower limits of the 95% confidence interval were 0.88, 0.82, and 0.77 at t_0, t_1, and t_2, respectively. As shown in Figure 7 b, in dataset B, the AUCs at t_0, t_1, t_2, t_3, and t_4 were 0.99, 0.76, 0.95, 0.89, and 0.93, respectively. The lower limits of the 95% confidence interval were 0.97, 0.56, 0.87, 0.74, and 0.83 at t_0, t_1, t_2, t_3, and t_4, respectively. In dataset B, the AUC and the lower limits of the 95% confidence interval of eENCD at t_1 were 0.76 and 0.56, which were lower than or equal to the other time points by the conventional method (Figure 8 and Figure 9). The AUCs and lower limits of the 95% confidence interval of eENCD were the highest in almost every case.

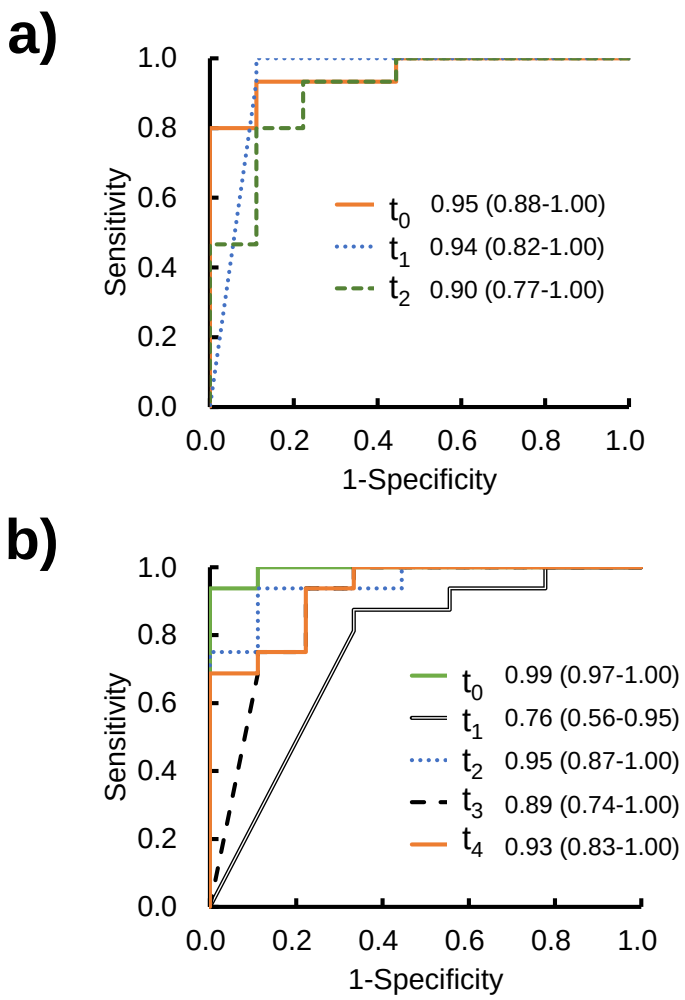

Figure 7 ROC curves of eENCD by bootstrap sampling. (a) ROC curve generated by eENCD at each time point of dataset A. The AUCs and 95% confidence interval are calculated by ROC curves at each time point. (b) ROC curve by eENCD at each time point of dataset B. As with (a), the AUCs and 95% confidence interval are calculated.

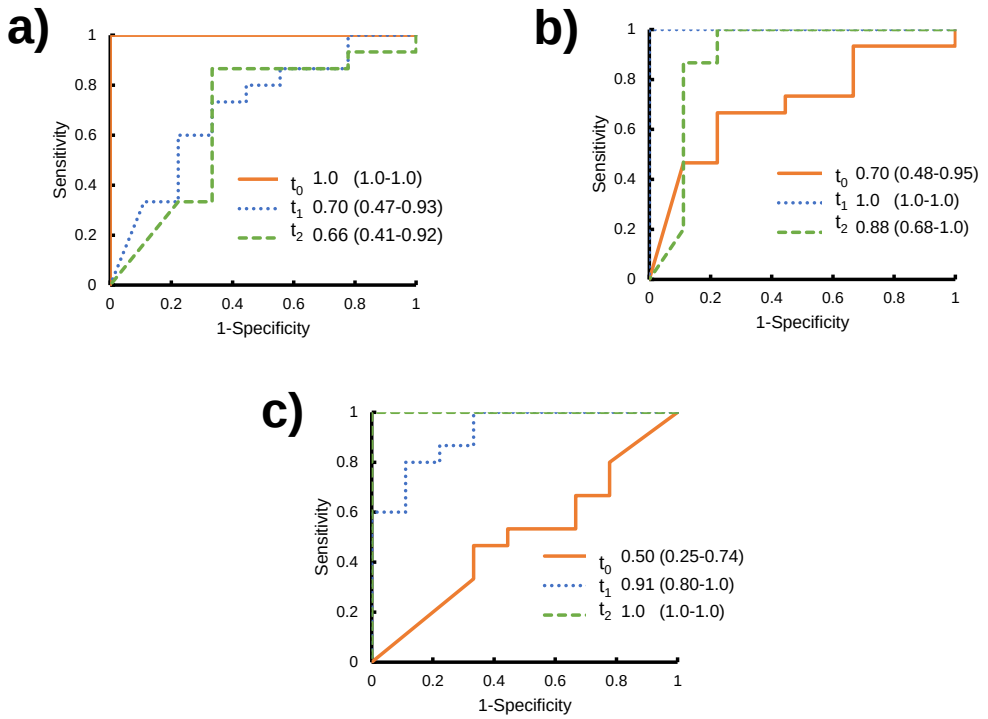

Figure 8 ROC of prediction model by the conventional method in dataset A. The AUCs and 95% confidence interval are calculated by ROC curves at each time point. Prediction model used t_0 (a), t_1 (b), and t_2 (c) data.

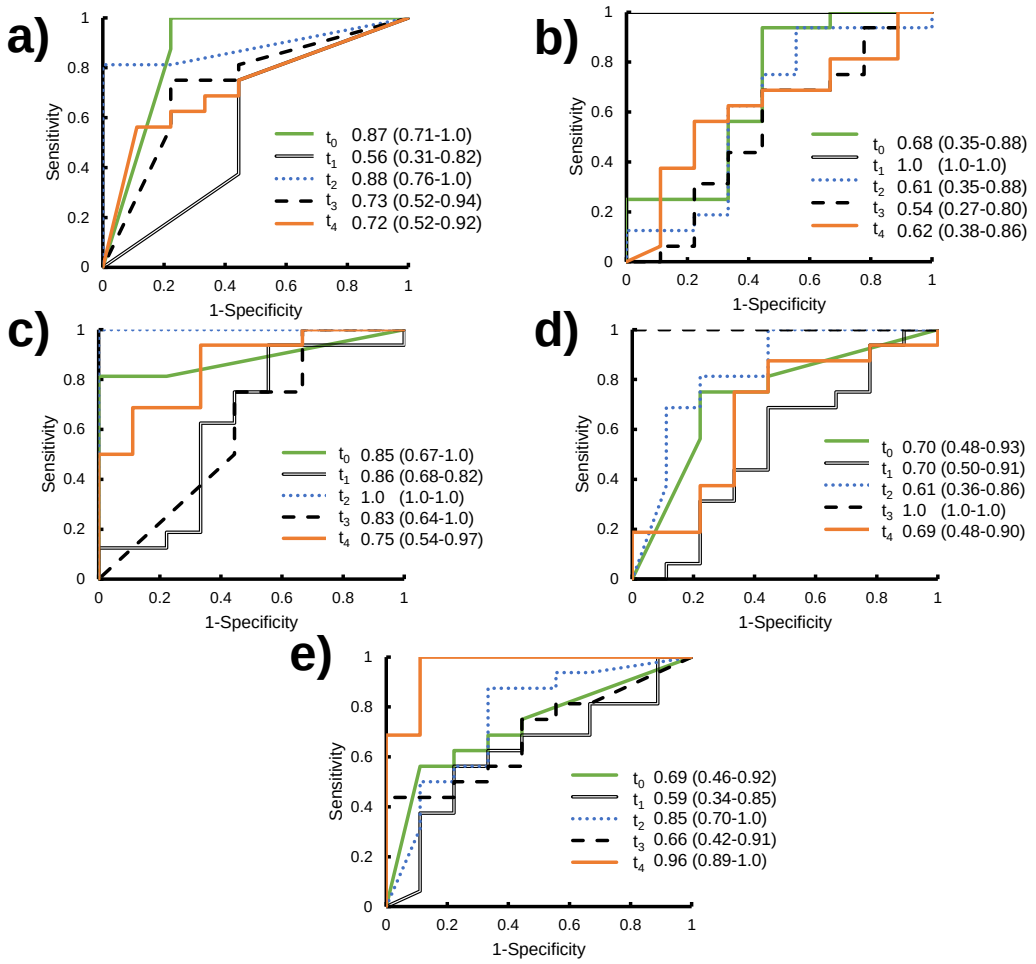

Figure 9 ROC of prediction model by the conventional method in dataset B. The AUCs and 95% confidence interval are calculated by ROC curves at each time point. Prediction model used at t_0 (a), t_1 (b), t_2 (c), t_3 (d), and t_4 (e) data.

Eleven genes were identified in dataset A by eENCD (Table 5) and eight genes were identified in dataset B by eENCD (Table 6). These genes were expected to have consistently higher expression levels of either sensitive or not sensitive responders at each time point. The median levels of 9 genes in dataset A were consistently differentiated throughout all time points (Table 5). In particular, the expression levels of the *HPS5* gene is not sensitive responders at t_0 and t_1 were significantly higher than those in sensitive responders ($p < 0.05$) (Figure 10 a). The median levels of 6 genes at each time point were consistently higher in either group in dataset B (Table 6). In particular, the expression levels of the *CDH2* gene of sensitive responders at t_0 and t_2 were significantly higher than those in not sensitive responders ($p < 0.05$) (Figure 10 b). From the above, eENCD identified some genes of which the expression levels were always consistently different throughout all time points.

Table 5 Identified genes of dataset A by eENCD.

Gene symbol	P value			Higher GE levels at all time points
	t_0	t_1	t_2	
ZBTB16	0.064	**0.013**	0.137	sensitive
ZFP37	0.070	0.220	**0.013**	--
HPS5	**0.013**	**0.013**	0.084	Not sensitive
HOPX	0.105	**0.005**	0.090	sensitive
ARFGAP3	**0.013**	0.162	0.105	sensitive
CALML5	0.077	**0.013**	0.126	sensitive
VPS26A	**0.026**	0.090	0.205	sensitive
SLC5A4	0.190	**0.022**	0.190	sensitive
MBL2	0.149	**0.013**	0.640	--
DLGAP4	**0.007**	0.115	0.390	sensitive
CACNA1C	0.064	0.382	0.390	Not sensitive

The p values are adjusted by BH method, and **Bold accuracy with underline** has difference gene expression (GE) levels between sensitive and not sensitive responder significantly ($p<0.05$). If GE levels of sensitive responders at each gene are higher than one of not sensitive at all time points, the sensitive is represented in the final column.

Table 6 Identified genes of dataset B by eENCD.

Gene symbol	P value					Higher GE levels of all TPs
	t_0	t_1	t_2	t_3	t_4	
SMA4	0.072	0.250	**0.009**	0.082	0.082	sensitive
MIR7114_NSMF	0.072	0.082	**0.005**	0.130	0.314	sensitive
LSM8	0.452	**0.009**	0.082	0.082	0.441	--
FLAD1	0.071	**0.009**	0.344	0.056	0.072	Not sensitive
RRN3P1	0.419	0.179	0.082	0.082	**0.023**	Not sensitive
RASL10A	**0.033**	0.334	0.344	0.452	0.314	--
IER3IP1	0.115	0.072	**0.005**	0.216	0.082	Not sensitive
CDH2	0.250	**0.033**	0.397	**0.043**	0.082	sensitive

The p values are adjusted by BH method, and **Bold accuracy with underline** has difference gene expression (GE) levels between sensitive and not sensitive responder significantly ($p<0.05$). If GE levels of sensitive responders at each gene are higher than one of not sensitive at all time points (TPs), the sensitive is represented in the final column.

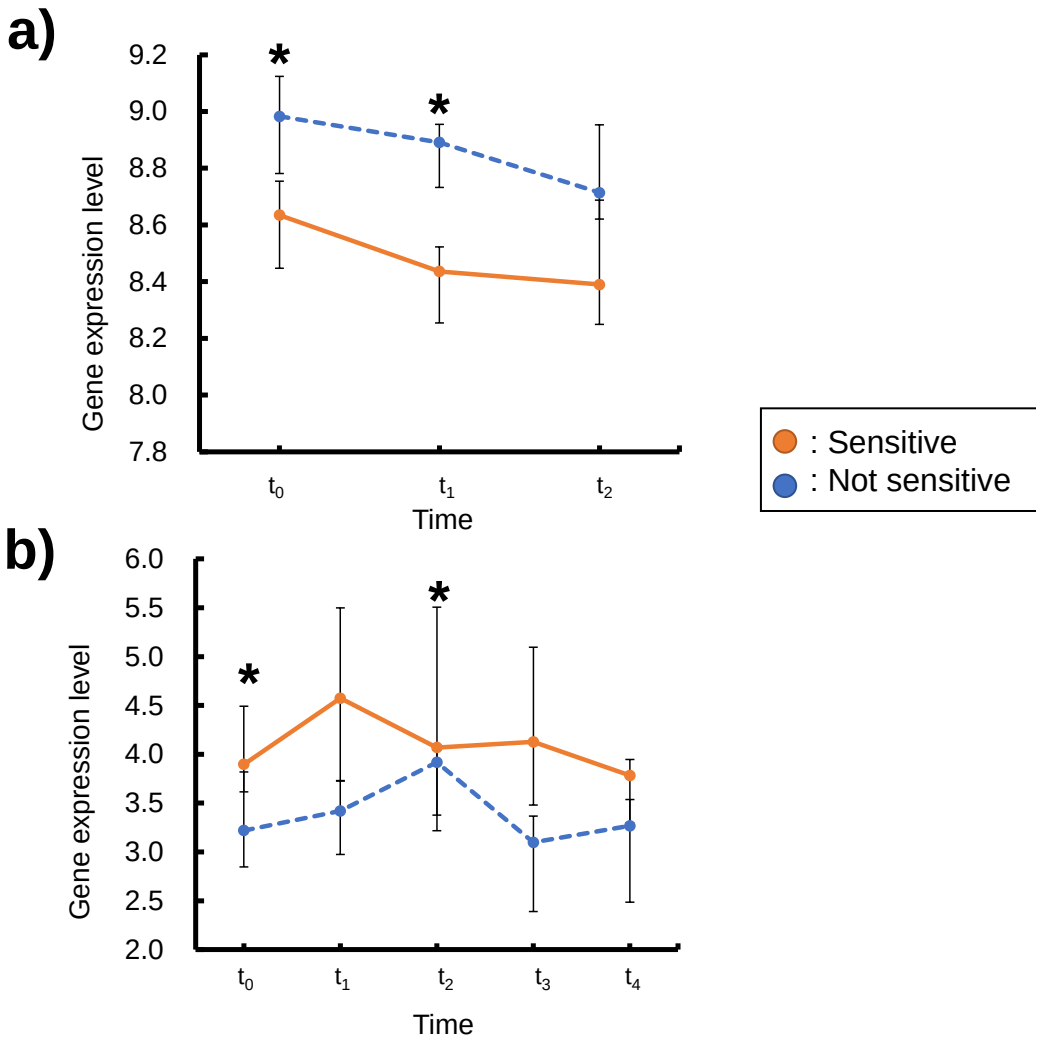

Figure 10 Gene expression levels of sensitive and not sensitive responders at each time point. Gene expression levels of the *HPS5* gene in dataset A (a) and the *CDH2* gene in dataset B (b). "*" is FDR-corrected $p < 0.05$ in the Wilcoxon rank-sum test.

2.4. Discussion of eENCD

The genes identified by eENCD showed consistent differentiation throughout all time points and accurately predicted the responses of MS patients to INF-β treatment.

The ACC_{mean} and ACC_{min} values by eENCD in dataset A were 86% and 79%, respectively. The ACC_{mean} value was equal to or higher than that by the conventional method (Table 3). The prediction model by the conventional method using t_1 data had an almost same ACC_{mean} value as eENCD; however, the ACC_{min} value of eENCD was higher than that of the conventional method. The ACC_{mean} and ACC_{min} values of eENCD in dataset B were 84% and 72%, respectively (Table 4). These values were higher than those of the conventional method. Thus, eENCD yielded higher and more accurate predictions throughout most time points than the conventional method (Table 3 and Table 4). In addition, the prediction accuracies at different time points were evaluated by bootstrap sampling. Figure 6 a and b show the mean and standard deviations of the prediction accuracies of different time points calculated by bootstrap sampling. The mean accuracy of different time points by eENCD was 81% and higher than those by the conventional method (Figure 6 a). This result indicates that eENCD could achieve significantly higher prediction accuracies than the conventional method at different time points. In dataset B, the mean accuracy of different time points was 78% and was significantly higher than those by the conventional method (Figure 6 b). Additionally, SES algorithm analysis of the static-longitudinal scenario is used as a conventional method [87], and this is compared with eENCD. The static-longitudinal scenario in this method is expected to identify genes showing consistent differentiation throughout all time points. Using the procedures of eENCD (Figure 6), the ACC_{mean} of the genes

Table 7 Selected genes and accuracy by bootstrap sampling using SES algorithm.

Dataset name	Dataset A	Dataset B
Identify gene symbol	*BID*	*CTDSPL*
time point of data for prediction model	t_0	t_4
Mean accuracy by bootstrap sampling [%]	57 ($p=2.32\times10^{-64}$)	61 ($p=2.92\times10^{-32}$)

The number in parentheses is a p-value of comparison between the SES algorithm and eENCD. For the implementation of the SES algorithm, MXM package was used in R.

identified by SES algorithm was calculated, and a prediction model was created. The prediction accuracies at different time points using SES algorithm were calculated by bootstrap sampling (Table 7). These mean accuracies obtained from our proposed method were higher than those given by this conventional method. Therefore, eENCD using time-course gene expression data could achieve a high prediction accuracy compared with those provided by the conventional methods. Given this, eENCD provided higher accuracy throughout all time points.

Figure 7 a and b show the sensitivity and specificity of eENCD; AUC was approximately 0.90 at most time points in both datasets A and B. However, the AUC at t_1 in dataset B by eENCD was 0.76, which was lower than the AUC at other time points and equivalent to the conventional method, as shown in Figure 8 and Figure 9. The results at each time point (Table 3, Table 4, Figure 11, and Figure 12) revealed that the prediction

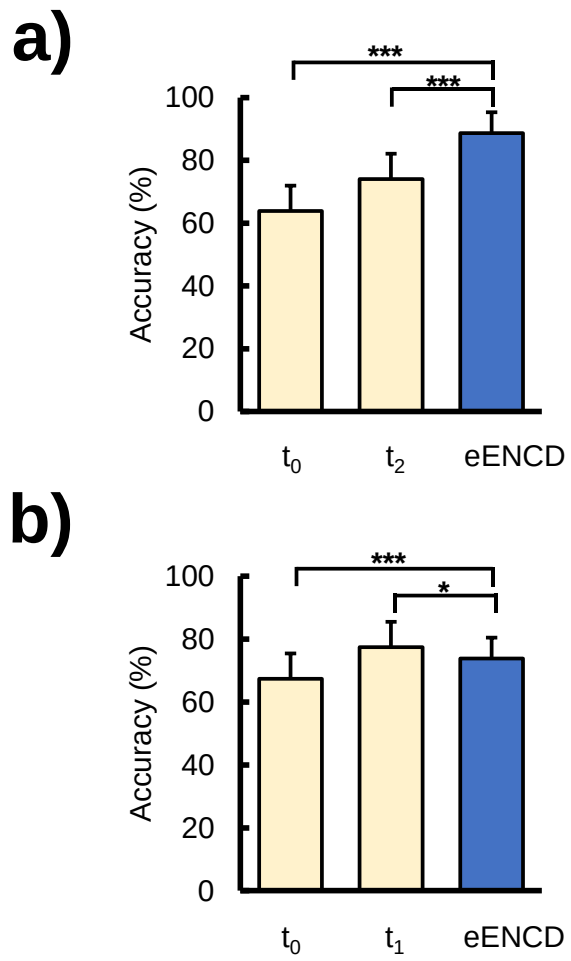

Figure 11 Prediction accuracies obtained using eENCD and conventional method at each time point using bootstrap sampling. The "*" and "***" mean that the p-values of the Wilcoxon rank-sum test are less than 0.05 and 0.001, respectively. Prediction accuracies of each model at t_1 (a) and t_2 (b) data in dataset A.

accuracies did not depend on the order of the time-course and that the prediction accuracies by eENCD were high at most time points; however, there was a case where the prediction accuracy was lower than that of the conventional method.

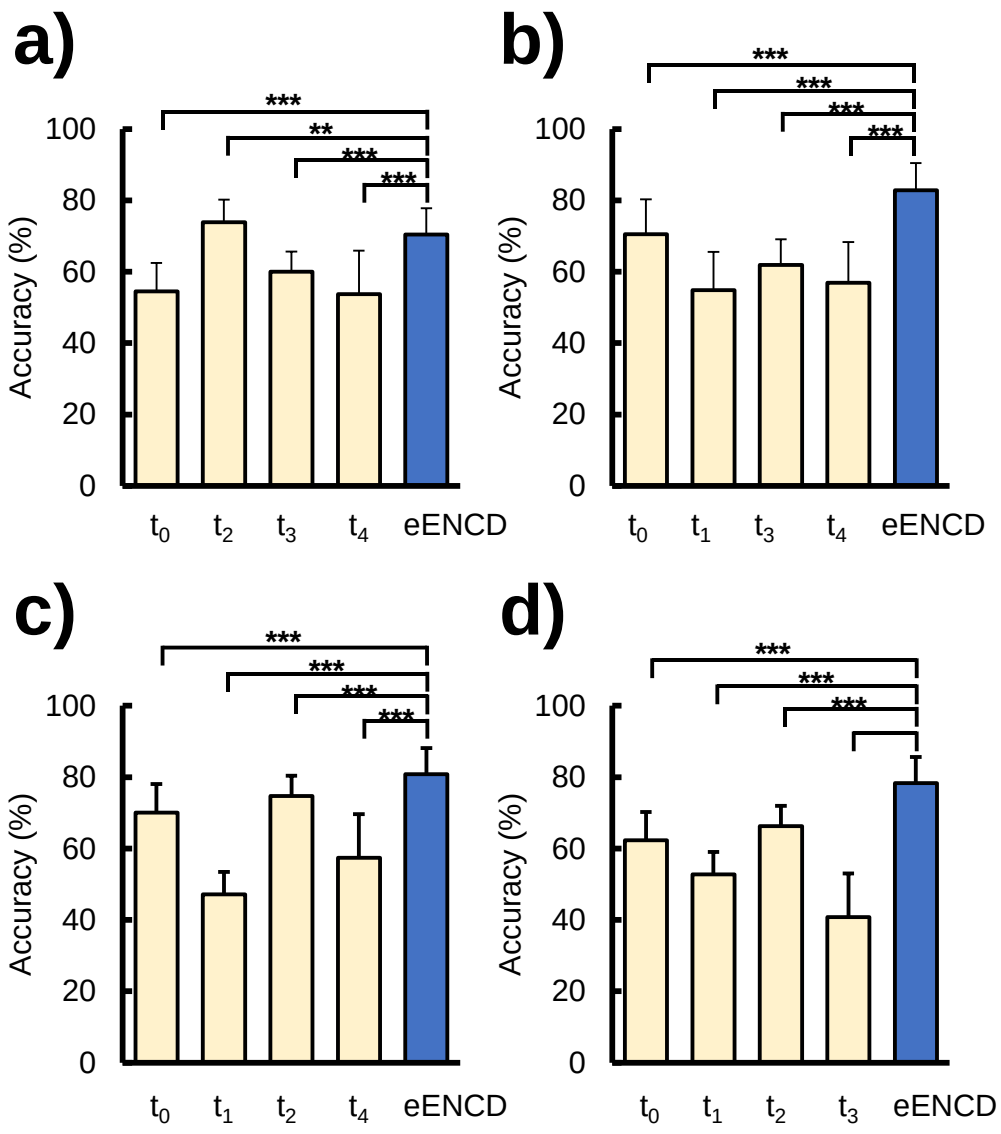

Figure 12 Prediction accuracies by eENCD and conventional method at each time point using bootstrap sampling. The "**" and "***" mean that the p-values of the Wilcoxon rank-sum test are less than 0.01, and 0.001, respectively.Predicton accuracies of each model at t_1 (a), t_2 (b), t_3 (c), and t_4 (d) data in dataset B.

As shown in Table 5 and Table 6, most genes of eENCD showed different gene expression levels consistently throughout all time points. Changes in those levels differentiated sensitive and not sensitive responders consistently throughout the time-courses significantly (Figure 10). From the above, eENCD identified genes showing consistent differentiation throughout multiple time points and could differentiate between sensitive and not sensitive responders.

The eENCD did not identify identical genes between datasets A and B. Regarding dataset A, associations between MS and *ZBTB16* and *HOPX* genes were reported [88]. *Th17* cells are a subset of T helper cells involved in many immune diseases including MS. *ZBTB16* gene was reported to activate differentiation of *Th17* cells, which contributed to the maintenance of the phenotype of *Th17* cells in the human body [88]. Regarding a relationship between the functional defect of *T* cells and autoimmune encephalomyelitis, many experiments and reviews reported the deletion of the *HOPX* gene for decreasing the suppressor ability of pTreg cells [89][90]. Regarding dataset B, there were reports on the *CDH2* gene [91]. Microglia, a type of glial cell of the central nervous system, is known as central immunocompetent cells, and the *CDH2 gene* is involved therein [91]. Many genes in dataset A were related to cancer but their association with MS was unclear.

The eENCD had several limitations. Firstly, there were time points where the prediction accuracy by eENCD was lower. Secondly, the γ value as the weights of the elastic net should be set adequately to ensure accurate prediction. Thirdly, we used genes as independent variables; however, the interactions of genes could also be considered as explanatory variables to obtain higher accurate predictions.

INF-β treatment is known to be effective in the prevention of relapses of MS; however, the accurate prediction of INF-β treatment responses is still necessary to solve the problem of individual variations and the side effects associated with treatment. Microarray has been used to identify genes for predicting therapeutic responses. However, there are several difficulties associated with microarray analysis, including high dimensionality and multicollinearity. In addition, the conventional method only allows an analysis of data at a single time point. Therefore, a novel method suitable for time-course data analysis should be developed. Here, we proposed eENCD to identify genes using the elastic net accommodating time-course data. The following are three features of eENCD: (1) sparse modeling was used to allow for the efficient identification of genes (gene numbers >> sample size); (2) the elastic net was used to prevent the multicollinearity of expression levels among genes; (3) the elastic net was modified to identify genes showing consistent differentiation throughout the time-course.

Two publicly available datasets were used to evaluate this method. The mean prediction accuracies of different time points were compared by the proposed and conventional methods. The accuracies obtained using two datasets were 71% and 73% for the conventional method, and 81% and 78%, significantly higher, for eENCD. The eENCD identified 11 and 8 genes in the two datasets. Differences in the expression levels of 9 and 6 genes between good and poor responders were consistent throughout the data at all time points. Therefore, the genes identified by eENCD were suggested to be capable of high-accuracy prediction throughout multiple time points. In addition, these genes included the genes reported to be related to MS by previous studies. The eENCD for the time-course data analyses was used to identify genes showing consistent differentiation between two

outcome groups throughout time-courses. Here, we demonstrated the use of this modified elastic net for the prediction of INF-β treatment responses in patients with MS. Moreover, this method could also be used for microarray time-course data analyses.

3. Consolidating probabilities of multiple time points: CPMTP

3.1. Background of CPMTP

Predicting a patient's response to therapy using various types of information is essential for designing systematic treatments [92][93]. HCV infection and MS are representative diseases showing individual variations that require personalized therapy. Systematic therapies utilizing pegylated interferon-alpha and ribavirin are recommended for the treatment of HCV infection [58]. However, only about half of all cases displayed a sustained response to this therapy [94]. Patients with HCV infection have reportedly exhibited serious neuropsychiatric side effects such as severe depression and psychosis [95]. INF-β is the most widely used MS therapy to control disease symptoms [69]. However, this therapy did not prevent almost half of all patients from relapsing and even developing symptoms of brain disease, as observed in some cases [68]. To make appropriate decisions regarding therapeutic strategies, such as cancellation or fixation of long-term therapy, the therapeutic response associated with these diseases must be accurately predicted via time-course monitoring [96][97]. Therefore, developing methods and markers that accurately predict individual therapeutic response is crucial for establishing successful long-term therapy.

Time-course gene expression profiling has advanced rapidly on account of time-course gene expression profiles collected from the same patient being more beneficial than those collected from the patient at a single time point [49][52]. Methods that determine gene

markers using time-course gene expression profiles are classified into two categories: statistical methods such as analysis of variance (ANOVA) [87][98] and machine learning such as sparse modeling [50], decision trees [58], clustering [34][99] and deep learning [100]. Many of these use standard problem settings to identify gene markers showing different time-course patterns between two groups, such as cases vs. control. Detecting different patterns in time-course gene expression profiles is extremely beneficial for clarifying the biological processes involved. However, sometimes it may cause difficulties in predicting therapeutic response. For example, gene markers indicating a massive change between two late-term therapy groups may pose a challenge when it comes to making an accurate prediction for the first term. Conversely, gene markers that indicate significant early-term changes in treatment may make accurate late-term prediction difficult. Therefore, gene markers that accurately predict response to therapy at each observed time point are preferable for predictive purposes.

In predicting a long-term therapeutic response, prediction accuracy may be improved by incorporating patient information, which is repeatedly observed for a marker over time [62][63][101]. Rizopoulos *et al.* [62] and Li *et al.* [63] proposed a new method that dynamically updates predictive indicators as time points increase; they suggested that their method may improve prediction accuracy. However, these markers were not gene markers but aortic gradient levels [62] and brain imaging indices [63], which were also clarified as being useful by other studies. Therefore, the current study assumed that using more time points to profile a gene marker would lead to more accurate therapeutic response predictions.

Here, we propose a new prediction model and a gene selection method using time-course gene expression profiles. This method is based on the hypothesis that improving the accuracy of predictions requires more information obtained from gene markers at multiple time points. Therefore, our prediction model was designed to consolidate information from multiple time points, and our gene selection method was designed to identify gene subsets as markers that predict therapeutic responses more accurately with increasing time points. Time-course microarray datasets collected from HCV and MS patients were used to evaluate the proposed method. In this evaluation, three types of experiments were performed as follows: (1) comparison with our proposed method and the conventional method; (2) hypothesis verification; and (3) function analysis of the gene subset selected by the proposed method.

3.2. Method of CPMTP

Our proposed method was designed to predict therapeutic response using multiple time-point data that would expectedly yield a higher level of accuracy than a prediction based on single-point data. Our method is termed the consolidating probabilities of multiple time points method. CPMTP consists of a prediction procedure and gene selection procedure. Sections 3.2.1 introduced the theory of CPMTPp and CPMTPg. Section 3.2.2 described the numerical experiments.

3.2.1. Theory for CPMTP

This section described CPMTPp and CPMTPg. Briefly, CPMTPp is the procedure for predicting therapeutic response using a model. CPMTPg is the procedure for selecting genes.

Concept of CPMTPp:

The CPMTPp design was based on the hypothesis that prediction accuracy is improved by consolidating information on the states of a patient at multiple time points. The general problem setting for the prediction in which the response at future time point T_{final} was estimated as either sensitive or not sensitive using gene expression profiles is shown (Figure 13). “Sensitive” meant that the patients responded well to therapy and recovered from the disease, and “not sensitive” meant that patients could not recover from the disease with the therapy. The time points corresponding to gene expression profiles used for prediction by CPMTPp and conventional methods were different. In this paper, the

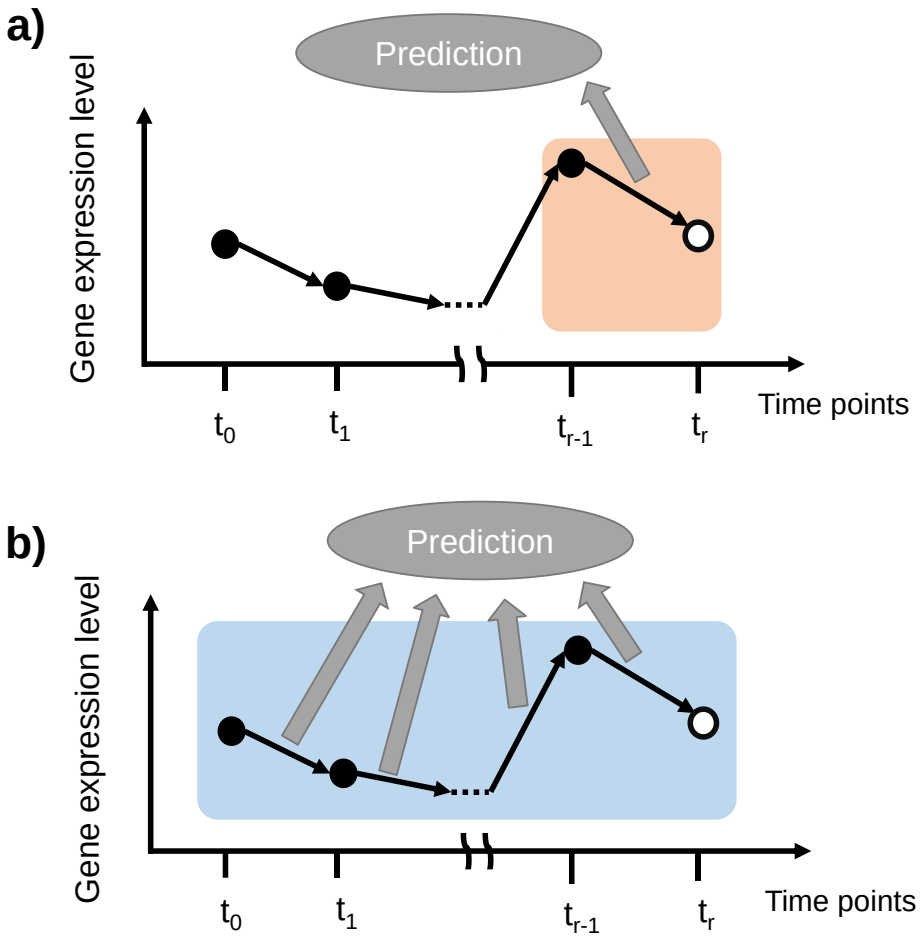

Figure 13 The concept of predicting therapeutic response using the conventional method and our proposed method. (a) The conventional method uses single gene expression levels or differential gene expression levels between two time points. (b) CPMTP uses time-course gene expression profiles according to consolidated probabilities of multiple time points.

time points used in each method are termed checkpoint (CP). The CP of conventional methods was a single-time point t_r $(r = 0, \dots, R)$ or the difference between the one-time

point t_r ($t_r \leq \mathrm{T}_{final}$) and the previous time point t_{r-1} (Figure 13 a); it was, thus, mostly confined to two-time points. Meanwhile, the CPMTPp used the gene expression profiles corresponding to the first time point t_0 to a time point t_r (Figure 13 b); here, more than two time points were used for predictive purposes. In this manner, the hypothesis was implemented using CPMTPp by consolidating the probabilities of therapeutic response using gene expression profiles collected from multiple time points.

CPMTPp was used to calculate one probability of therapeutic response using time-course microarray data (Figure 14). Firstly, CPMTPp was used to calculate a probability using the gene expression profile at a time point, t_r. Secondly, the probability at t_r and the prior probability from t_0 to t_{r-1} was consolidated to calculate a more accurate probability. By repeating these two steps until $\mathrm{r} = R$, the probabilities at multiple time points were aggregated into one probability (where T was the final time point that can be used for prediction).

Similar to the Bayesian model [102][103] and neural network [104], multiple logistic regression (MLR) models have been widely used to predict the response to therapy based on probability. This probability did not present a p-value in statistical tests but present how likely the patient is likely sensitive (or not sensitive). The probability at the first step $\mathrm{P}_{logit}^{(t_r)}$ was calculated through MLR, like Eq. 7, using the difference of gene expression profile between t_r and t_{r-1} (Eq. 8). However, these models used single or two time points to calculate the probability and did not use time-course data.

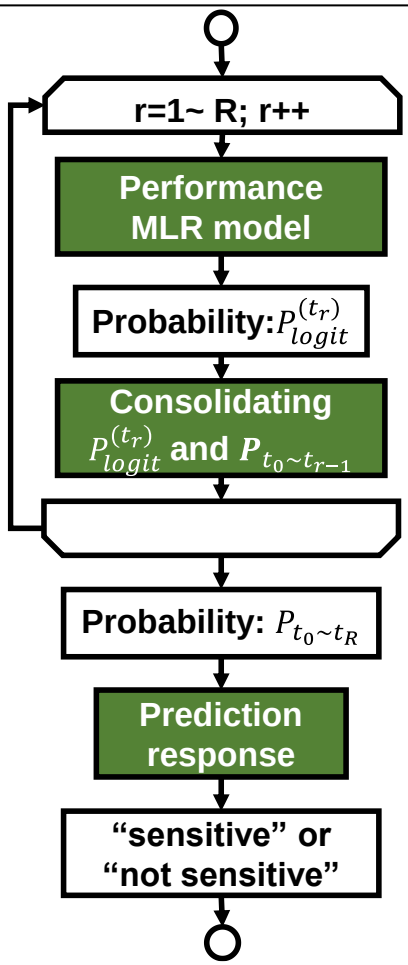

Figure 14 The flow of CPMTPp. CPMTPp is based on the hypothesis that information at multiple time points improves prediction accuracy.

$$\mathrm{P}_{logit}^{(t_r)} = \frac{1}{1 + e^{-(w_0^{(t_r)} + w_1^{(t_r)} * d_1^{(t_r)} + \cdots + w_l^{(t_r)} * d_l^{(t_r)})}}$$

Eq. 7

where

$d_j^{(t_r)}$ $(j = 1, \dots, l)$: different expression levels of the j^{th} gene between two-time points t_r and t_{r-1}. l is the number of genes in a gene subset of logistic regression. The gene subset was selected from all the genes collected by microarray using CPMTPg.

$w_j^{(t)}$ $(j = 0, \dots, l)$: the weight of j^{th} gene as a feature in a gene subset. $w_0^{(t_r)}$ is a constant term at time point t_r.

$$d_j^{(t_r)} = x_j^{(t_r)} - x_j^{(t_{r-1})}$$

Eq. 8

where

$x_j^{(t_r)}$: j^{th} gene expression levels at time point t_r. $x_j^{(t_{r-1})}$ is the j^{th} gene expression level at time point t_{r-1}.

In CPMTPp, the Bayesian theory was used to consolidate probabilities based on time-course data [68]. The probability $P_{t_0 \sim t_1}$ was calculated by combining the probability at time point $P_{logit}^{(t_r)}$ and the probability at previous time points $P_{t_0 \sim t_{r-1}}$ (Eq. 9). As the previous time point did not exist ($r = 1$), $P_{t_0 \sim t_1} = P_{logit}^{(t_1)}$ was defined. $P_{t_0 \sim t_r} \geq 0.5$ and $P_{t_0 \sim t_r} < 0.5$ indicate sensitive and not sensitive responses, respectively (Eq. 10). From the above, CPMTPp could be used to predict response to therapy based on gene expression profiles at multiple time points.

$$\mathrm{P}_{t_0 \sim t_r} = \frac{\mathrm{P}_{logit}^{(t_r)} * \mathrm{P}_{t_0 \sim t_{r-1}}}{\mathrm{P}_{logit}^{(t_r)} * \mathrm{P}_{t_0 \sim t_{r-1}} + (1 - \mathrm{P}_{logit}^{(t_r)}) * (1 - \mathrm{P}_{t_0 \sim t_{r-1}})}$$

Eq. 9

where

$\mathrm{P}_{logit}^{(t_r)}\ (t_r = t_0, \ldots, t_R)$: A probability of sensitive (or not sensitive) response to therapy using gene expression profile at time point t_r.

$$Predicted\ therapeutic\ response \coloneqq \begin{cases} sensitive & (\mathrm{P}_{t_0 \sim t_r} \geq 0.5) \\ not\ sensitive & (\mathrm{P}_{t_0 \sim t_r} < 0.5) \end{cases}$$

Eq. 10

where

$\mathrm{P}_{t_0 \sim t_r}\ (t_r = t_0, \ldots, t_R)$: A probability of sensitive (or not sensitive) response to therapy using gene expression profile at time points $t_0 \sim t_R$.

Algorithm of CPMTPg:

CPMTPg were used to select the gene subset of CPMTPp best suited for accurate prediction using time-course microarray data. CPMTPg was used to decide the gene subset by optimizing the fitness function based on the probability $\mathrm{P}_{t_0 \sim t_r}$ used in CPMTPp. This function was designed with negative penalties for incorrect predictions. The CPMTPg flowchart, which consists of gene screening (step 1) and deciding on a gene subset (step 2), is shown (Figure 15).

Step 1: The elastic net with the stability selection eliminated genes with low impact on therapeutic responses, yielding a gene pool.

Step 2: The gene subset was selected from the gene pool via an optimization method.

Here, the gene expression profiles were composed as a data matrix (p genes × n subjects × R time points). Each subject was labeled as sensitive or not sensitive based on therapeutic responses.

[Step1: Screening step]

Gene selection based on microarray data frequently suffers from the $n >> p$ problem, i.e., a large number of genes (p) compared to the small number of samples (n) [105]. Gene selection using univariate analyses causes an α-error by independent multiple tests. These p-values should be corrected via adjusting using methods such as the Bonferroni correction [106], Holm method, or Dunnett's method [98]. However, sparse modeling enables the selection of genes without p-values.

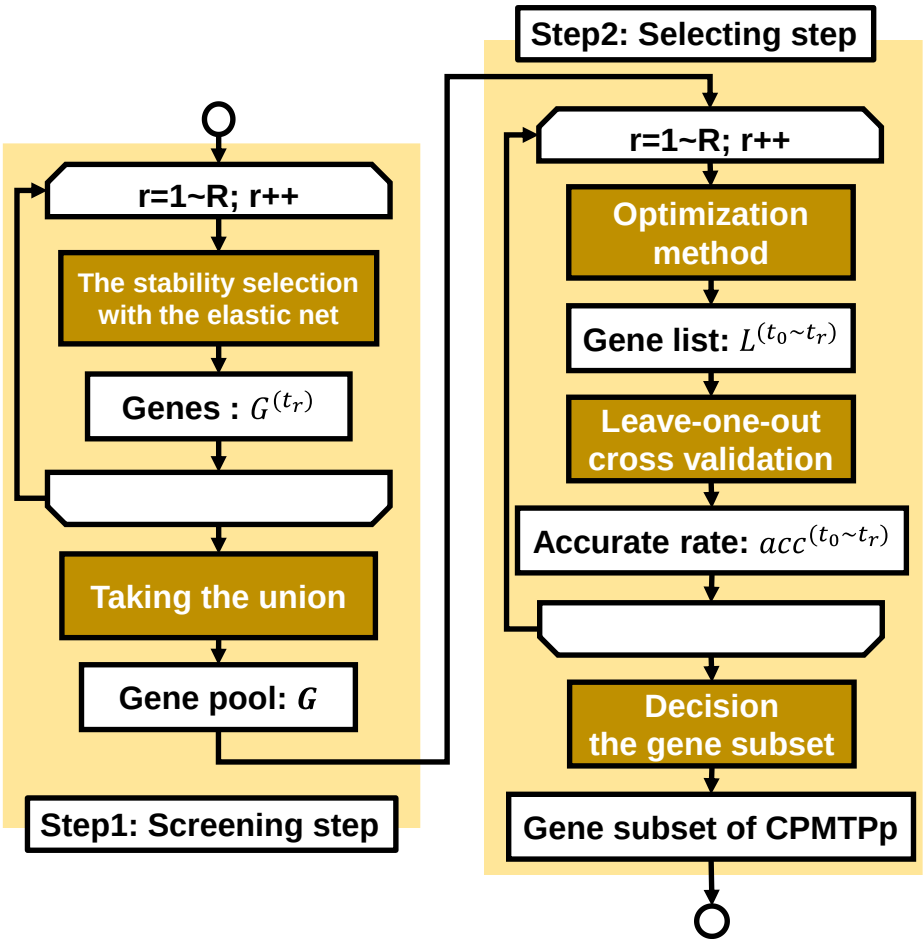

Figure 15 The flow of CPMTPg. CPMTPg selects a gene subset used by the CPMTPp model.

The sparse modeling solved the $p >> n$ problem by considering a condition where only a few genes affect the phenomenon under focus [34][50][107]. We employed the elastic net, a sparse modeling method [47]. The elastic net selects a subset effectively from features with high multicollinearity. To eliminate genes with minimal impact on therapeutic responses, the elastic net was applied to gene expression data at each time point.

The elastic net was used to select genes with non-zero weights, $w_j^{(t_r)}$, which are used for an MLR model according to Eq. 11. The genes with zero weights indicate that these genes were not selected as a gene subset for the MLR model. The elastic net equation added regularized terms (the second and third terms of Eq. 11) to the general loss function, such as the least-squares method (the first term of Eq. 11) to optimize the weights. The second term prevents multicollinearity, and the third term selected features. Because the genes selected by the elastic net depended on the value of lambda in Eq. 11, deciding lambda was important to predict the response to therapy accurately.

$$\underset{\boldsymbol{w}^{(t)}}{\operatorname{argmin}} J\left(\boldsymbol{X}^{(t_r)}, \boldsymbol{y}\right) + \lambda\left(\sum_{j=1}^{l}(1-\alpha)\frac{1}{2}\left|w_j^{(t_r)}\right|^2 + \sum_{j=1}^{l}\alpha\left|w_j^{(t_r)}\right|\right)$$

Eq. 11

where

$\boldsymbol{w}^{(t_r)} = (w_1^{(t_r)}, \ldots, w_l^{(t_r)})^{\top}$: the weights of a logistic regression model at a time point t_r in Eq. 10.

$\boldsymbol{X}^{(t_r)} = \left(\boldsymbol{x}_1^{(t_r)}, \ldots, \boldsymbol{x}_p^{(t_r)}\right)$; $\boldsymbol{x}_j^{(t_r)} = \left(x_j^{(1,t_r)}, \ldots, x_j^{(N,t_r)}\right)^{\top}$: the difference in j^{th} gene expression levels at a time point between t_r and t_{r-1}.

$\boldsymbol{y} = (y^1, \ldots, y^n)$; $y^i \in \{0,1\}$: Therapeutic response of i^{th} patient at time point t. If $y^i = 1$, the therapeutic response presents "sensitive." If not, the therapeutic response presents "not sensitive."

$J\left(\boldsymbol{X}^{(t_r)}, \boldsymbol{y}\right)$: Loss function of a logistic regression model using $\boldsymbol{X}^{(t_r)}$ and $\boldsymbol{y}$.

λ: A hyper-parameter that represents the weight of the regularized terms in Elastic Net.

α ($0 < \alpha < 1$): A hyper-parameter that decides the assignment of the second and third terms.

The stability selection was used to reduce the effect of lambda on feature selection [33][83]. The stability selection performed the elastic net many times with various lambda values to sub-sample sets via random sampling. A gene pool at each time point ($\boldsymbol{G}^{(t_r)}$ in Figure 15) was created based on the selected rate in repeated times at a lambda value. At step 1, the gene list ($\boldsymbol{G}$ in Figure 15) consisted of genes belonging to gene pools at any of the time points $\{\boldsymbol{G}^{(t_0)}, \ldots, \boldsymbol{G}^{(t_R)}\}$. In this step, some genes that affected prediction at each time point could be selected from a huge number of genes in the microarray data.

[Step2: Selecting a gene subset]

In step 2 of CPMTPg, the gene subset for the CPMTPp model was selected from the gene pool $\boldsymbol{G}$ via optimization. For CPMTPg, the gene list ($\boldsymbol{L}^{(t_0 \sim t_r)}$ in Figure 15) was created by combinatorial optimization method.

This step was performed as follows:

(i) The gene list ($\boldsymbol{L}^{(t_0 \sim t_r)}$ in Figure 15) was selected from gene expression profiles at time points from t_0 to t_r via the optimization method.

(ii) The subjects in the gene expression data were separated into two blocks.

(iii) The CPMTPp model was constructed based on one block of data using the gene list $\boldsymbol{L}^{(t_0 \sim t_r)}$.

(iv) The accurate rate of the model ($acc^{(t_0 \sim t_r)}$ in Figure 15) was calculated using the other block of data.

(v) (ii) and (iv) were repeated for r =R.

(vi) The gene list showing the best accuracy rate was determined as the gene subset of CPMTPp.

The fitness function of the optimization method was designed via probability consolidated at multiple time points (Eq. 12). Eq. 12 used the probability, $P_{t_0 \sim t_r}^{(s)}$, and the number of accurate predictions in patients, N_{true}, as a reward, and the probability, $P_{t_0 \sim t_r}^{(q)}$, and the number of patients with no incorrect predictions, N_{False}, as a penalty. The absolute value of the difference between probabilities and 0.5 ($\left|P_{t_0 \sim t_r}^{(s)} - 0.5\right|$ and $\left|P_{t_0 \sim t_r}^{(q)} - 0.5\right|$ in Eq. 12) presented a confidence level of the predicted therapeutic response. If probabilities $P_{t_0 \sim t_r}^{(s)}$ and $P_{t_0 \sim t_r}^{(q)}$ were closer to 0 or 1, respectively, these values were higher. However, if these probabilities were closer to 0.5, these values were lower. Therefore, the first and second terms of Eq. 12 are the mean values of the confidence levels of accurate and incorrect predicted therapeutic responses, respectively. The optimization method selects the gene subsets with a CPMTPp that can accurately predict and display high confidence levels for the predicted response to therapy by maximizing the fitness

function of Eq. 12.

$$Fitness := \frac{1}{N_{true}} \sum_{s=1}^{N_{true}} \left| P_{t_0 \sim t_r}^{(s)} - 0.5 \right| - \frac{1}{N_{false}} \sum_{q=1}^{N_{false}} \left| P_{t_0 \sim t_r}^{(q)} - 0.5 \right|$$

Eq. 12

where

N_{true}: Number of patients in whom the actual therapeutic response equaled the predicted one.

N_{false}: Number of patients in whom the actual therapeutic response did not equal the predicted one.

$P_{t_0 \sim t_r}^{(s)}$ $(\mathrm{s} = 1, \ldots, N_{true})$: Probability of s^{th} patients that the actual therapeutic response equaled the predicted one.

$P_{t_0 \sim t_r}^{(q)}$ $(\mathrm{q} = 1, \ldots, N_{false})$: Probability of q^{th} patients that the actual therapeutic response did not equal the predicted one.

To determine the gene subset of CPMTPp from gene lists $\{\boldsymbol{L}^{(t_0 \sim t_1)}, \ldots, \boldsymbol{L}^{(t_0 \sim t_R)}\}$, the accurate rate $acc^{(t_0 \sim t_r)}$ was calculated by the (ii)–(iv) flows. In these flows, leave-one-out cross-validation was used. The number of patients in two blocks of data by this cross-validation was 1 for evaluation and $N - 1$ for the construction of the model. The accuracy rate was shown as the proportion of patients whose predicted therapeutic responses were accurate for evaluating the cross-validation.

CPMTPg made it possible to construct a CPMTPp model that enabled accurate prediction at multiple time points using the gene list with the highest accuracy rate as the gene subset in CPMTPp.

Note that this step used the ridge as an optimization method for weights $w^{(t_r)}$ in an MLR model (Eq. 7) to calculate $P^{(s)}_{t_0 \sim t_r}$ and $P^{(q)}_{t_0 \sim t_r}$ inEq. 9. The ridge does not select genes and constructs the model to avoid multicollinearity. At $\alpha = 0$ of Eq. 11, this equation is not the elastic net, but the ridge.

3.2.2. Numerical Experiments for CPMTP

Three experiments were performed: (1) comparison with CPMTP (CPMTPp+CPMTPg) and a conventional method, (2) verification of our hypothesis, and (3) analysis of the gene subset selected by CPMTPg. This section describes the material, preprocessing, evaluation method, parameters, and implementation.

Material and Preprocessing:

Two sets of time-course microarray data were used for this evaluation. One dataset was collected from HCV patients treated with antiviral therapies, peginterferon, and ribavirin (HCV dataset) [108]. The other was collected from MS patients treated with INF-β (MS dataset) [71]. These datasets (GSE7123 and GSE24427) were opened on the GEO website (https://www.ncbi.nlm.nih.gov/geo/).

The details of these datasets are shown (Table 8). The number of time points in the HCV dataset was six (t_0 to t_5), and the difference in CP between CPMTPp and a conventional method is shown (Table 9). There were five MS datasets (t_0 to t_4), and the difference in CP between CPMTP and a conventional method is shown (Table 10). Gene expression profiles were collected using the Affymetrix Human Genome U133A Array from peripheral blood mononuclear cells of patients, where the patients used for this evaluation were limited to those who could provide these at all time points.

Table 8 Summary of microarray data collected from HCV and MS patients.

Dataset name	HCV dataset	MS dataset
Number of genes	13513	13513
Number of time points	0 day(t_0), 1 day(t_1), 2 days(t_2), 7 days(t_3), 14 days(t_4), 28 days(t_5)	first(t_0), second(t_1), 1 month(t_2), 12 months(t_3), 24months(t_4)
Number of sensitive/not sensitive responders	sensitive: 36 not sensitive: 22	sensitive: 16 not sensitive: 9
Number of sensitive/not sensitive responders for stability selection	sensitive: 28 not sensitive: 17	sensitive: 12 not sensitive: 7
Number of sensitive/not sensitive responders for *k*-fold cross-validation	sensitive: 12 not sensitive: 7~8	sensitive: 10~11 not sensitive: 6

The number of genes of both HCV and MS is 13513, but the types of genes are different from them. In this paper, symbols of time points are presented as t_0, t_1, t_2 etc.

Table 9 The difference in CPs between the conventional method and CPMTP in the HCV dataset.

Time point	CP1	CP2	CP3	CP4	CP5
t_0	# *	*	*	*	*
t_1	# *	# *	*	*	*
t_2		# *	# *	*	*
t_3			# *	# *	*
t_4				# *	# *
t_5					# *

Table 10 The difference in CPs between the conventional method and CPMTP in the MS dataset.

Time point	CP1	CP2	CP3	CP4
t_0	# *	*	*	*
t_1	# *	# *	*	*
t_2		# *	# *	*
t_3			# *	# *
t_4				# *

"#" and "*" are the time points of microarray data used by the conventional method and CPMTP, respectively.

Three steps were performed to preprocess gene expression data. Several probes were removed from the two datasets. As the probes had duplicate gene symbols in one dataset, one probe was selected by comparing median gene expression levels, and the other probes were removed. Probes with a gene symbol indicating a non-coding region or no gene symbol were also removed. After removing these probes, $\log_2$ fold-change and quantile normalization were performed on each dataset.

Conventional method:

The MLR model was used as the prediction model for the conventional method. The features of the MLR model were based on the difference in gene expression profiles between t_r and t_{r-1}. The CPs of the MLR model for the HCV and MS datasets are shown in Table 9 and Table 10, respectively.

Next, maSigPro was used for gene selection in the conventional method, which is frequently used for time-course microarray data analysis [106][109][110]. This method selects the gene subset which shows a time-course difference in gene expression profiles between two groups via p-values of a statistical test with the significant level of $s_{maSigPro}$. This p-value was associated with F-statistic and was corrected by the linear step-up false discovery rate procedure. When the number of genes selected by maSigPro was over l_{max}, l_{max} genes were selected in ascending order of p-values.

Evaluation method:

To compare CPMTPp+CPMTPg and MLR+maSigPro as the conventional method, the AUC and accuracy were calculated using HCV and MS datasets. For this, k-fold cross-

validation was performed. This method splits patients in the dataset into k blocks. The k-1 blocks were used for the model training, and the remaining 1 block was used for evaluation. This procedure was repeated k times, and all data were used for evaluation at one time.

The ROCs for each CP, which were calculated based on probabilities of CPMTPp+CPMTPg and MLR+maSigPro that were obtained via k-fold cross-validation, are depicted. The AUCs were calculated using these ROC curves. The difference between AUCs corresponding to CPMTPp+CPMTPg and MLR+maSigPro at each CP were compared using the DeLong test with significance levels s_{AUC}.

To compare with CPMTP and previous studies based on therapeutic responses estimated via k-fold cross-validation, the accuracies of CPMTPp+CPMTPg and MLR+maSigPro were calculated. The accuracies were calculated for each CP and each block for evaluation in k-fold cross-validation. Based on the mean, maximum, and minimum values of these accuracies, CPMTPp+CPMTPg and MLR+maSigPro were compared.

In CPMTPp, it was assumed that the accuracy of the prediction model was improved as time points increased. The accuracies of the CPMTPp and MLR models were compared to verify this hypothesis. The gene selection methods of these models were CPMTPg. The mean, maximum, and minimum values of accuracies in CPMTPp+CPMTPg and MLR+CPMTPg were calculated using k-fold cross-validation using HCV and MS datasets.

The gene subset selected by CPMTPg was analyzed by ontology to research the function of genes in the biological process. DAVID (https://david.ncifcrf.gov/home.jsp)

was used as an ontology analysis tool. Common terms (GO terms) that were associated with the genes of CPMTPg were decided using DAVID based on p-values below the significance level s_{DAVID}. This p-value was a modified Fisher exact p-value. Also, previous studies were investigated using GO terms as keywords.

Parameters and Implementation:

The therapeutic responses of patients were decided at the final time point of the datasets. The final time point in the HCV dataset was 28 days (t_5) after the first therapy. In the HCV dataset, the types of therapeutic responses were categorized as "marked," "intermediate," and "poor". The marked response was defined as "decreasing RNA levels of HCV $>$ 3.5 log10 IU/ml" or no detected levels on day 28. The intermediate response was defined as "decreasing 1.4 $\leq$ RNA levels of HCV $\leq$ 3.5 log10 IU/ml " on day 28. The poor response was defined as " decreasing RNA levels of HCV $<$ 1.4 log10 IU/ml " on day 28. However, for this evaluation, the marked and intermediate responses were considered as good results, as in previous studies [58]. Responses to therapy based on the MS dataset were decided by the occurrence of relapse up to 24 months (t_4) after first therapy, and they were considered "good" or "poor". In this paper, a "good" response was treated as "sensitive," and a "poor" response was treated as "not sensitive."

The parameters of Step 1 in the CPMTPg are as follows. The stability selection was repeated 100 times. The stability selection selected 80% of patients from each sensitive and not sensitive category as the sub-sample set (Table 8). Lambda values corresponding

to repetition were created based on the exponential function from $log10^{-3}$ to $log3$. The alpha value of the elastic net in the stability selection was 0.5.

The parameters of Step 2 in the CPMTPg were as follows. A genetic algorithm (GA) was utilized as the optimization method. GA is a heuristic optimization method that has been frequently utilized as a gene selection method for microarray data [105][111][112]. GA repeats single-point crossover, ranking selection, and mutation at each generation. The number of generations was 50, and the population size of each generation was 20. The phenotype of GA is a binary presented as either to select or not select gene candidates. Note that the maximum number of genes selected for each population was 10 and that the population for the first generation was created by random sampling. To create the next generation, a single-point crossover was generated twice in the population, and mutation was performed on 20% of the population. The mutation reversed the select or not select process at a randomly chosen locus in the population. Based on the fitness values in Eq. 6, ranking selection identified the top 40% and the bottom 10% of the total population as the next generation.

The parameters of the numeric experiment are as follows: $k = 3$ in *k*-fold cross-validation, and the rate of patients whose therapeutic responses were sensitive or not sensitive was the same for all blocks (Table 8). The parameters of maSigpro were $s_{maSigPro} = 0.05$ and $l_{max} = 10$. The significance level of the DeLong test and DAVID were $s_{AUC} = 0.05$ and $s_{DAVID} = 0.05,$ respectively.

The implementation language was R-Language (ver. 3.6.0). Quantile normalization, the elastic net, and maSigPro were used by "*limma (ver. 3.40.6)*," "*glmnet (ver. 2.0-18)*," and "*maSigPro (ver. 1.56.0)*" packages, respectively. The stability selection and GA were

implemented by the authors. The source codes used in this paper will be made available upon request. The pseudo-code of CPMTP was added in Figure 16, Figure 17, and Figure 18.

Algorithm: CPMTPp

Input: $\boldsymbol{x} = \{\boldsymbol{x}^{(t_1)}, \ldots, \boldsymbol{x}^{(t_R)}\}, \boldsymbol{w} = \{\boldsymbol{w}^{(t_1)}, \ldots, \boldsymbol{w}^{(t_R)}\}$

Output: $\hat{y} \in \{1: sensitibe, 0: not\ sentitive\}$

01: **for** r = 2,...,R **do**

02: Calculate $\boldsymbol{d}^{(t_r)} = \{d_1^{(t_r)}, \ldots, d_l^{(t_r)}\}$ using $\boldsymbol{x}^{(t_r)}$ (Eq.2)

03: Calculate $P_{logit}^{(t_r)}$ using $\boldsymbol{w}^{(t_r)}$and $\boldsymbol{d}^{(t_r)}$ (Eq.1)

04: **if** r==2 **do**

05: $P_{t_0 \sim t_r} = P_{logit}^{(t_r)}$

06: **else then**

07: Calculate $P_{t_0 \sim t_r}$ using $P_{t_0 \sim t_{r-1}}$ and $P_{logit}^{(t_r)}$ (Eq. 3)

08: **end if**

09: **end for**

10: Decide $\hat{y}$ using $P_{t_0 \sim t_r}$ (Eq. 4)

Figure 16 The pseudo-code of CPMTPp. This code predicts a therapeutic response of a patient. $\boldsymbol{x}^{(t_r)} = (x_1^{(t_r)}, \ldots, x_l^{(t_r)})$ $(r = 1, \ldots, R)$: gene expression levels collected by the patient at time point t_r. l is the number of genes in the gene subset. $\boldsymbol{w}^{(t_r)} = (w_1^{(t_r)}, \ldots, w_l^{(t_r)})^\top$: wights of the logistic regression at time point t_r. $w_0^{t_r}$ is a constant term. $\hat{y}$ is the predicted sensitive or not sensitive of the patient.

Algorithm: CPMTPg: Screening

Input: $\boldsymbol{X} = \{\boldsymbol{X}^{(t_0)}, \dots, \boldsymbol{X}^{(t_R)}\}, \boldsymbol{y} = \{y^{(1)}, \dots, y^{(n)}\}, \boldsymbol{\lambda} = \{\lambda_1, \dots, \lambda_K\}$

Output: Gene pool $\boldsymbol{G} \in \{1, \dots, p\}$

01: **for** r = 1,...,R **do**
02: **for** k= 1,...,K **do**
03: **for** s = 1,...,K **do**
04: Create a sub-sample set from $\{1, \dots, n\}$
05: Create $\boldsymbol{X}_{sub}$ and $\boldsymbol{y}_{sub}$ from $\boldsymbol{X}$ and $\boldsymbol{y}$ by selecting samples in the sub-sample set
06: $\boldsymbol{D}_{sub}^{(t_r)} = \boldsymbol{X}_{sub}^{(t_r)} - \boldsymbol{X}_{sub}^{(t_{r-1})}$
07: Optimize $\boldsymbol{w}^{(t_r)}$ using $\boldsymbol{D}_{sub}^{(t_r)}$, $\boldsymbol{y}_{sub}$, and λ_k (the elastic Net: Eq. 5)
08: Calculate $\boldsymbol{\psi}^{(s)} = \left\{\psi_1^{(s)}, \dots, \psi_p^{(s)}\right\}$ using $\psi_j^{(s)} = \left\{1 \; if \; w_j^{(t_r)} \neq 0; \; 0 \; if \; w_j^{(t_r)} = 0\right\} (j = 1, \dots, p)$
09: **for end**
10: Calculate $\boldsymbol{\pi}^{(k)} = \left\{\pi_1^{(k)}, \dots, \pi_p^{(k)}\right\}$ using $\pi_j^{(k)} = (\frac{1}{S_{repeat}}) \sum_{s=1}^{S_{repeat}} \psi_j^{(s)}$
11: **for end**
12: Calculate $G^{(t_r)} \in \{1, \dots, p\}$ by select genes whose maximum value of $\left\{\pi_j^{(1)}, \dots, \pi_j^{(K)}\right\}$ is more than 0.5.
13: **for end**
14: Create $\boldsymbol{G}$ by taking the union $\left\{G^{(t_1)}, \dots, G^{(t_R)}\right\}$

Figure 17 The pseudo-code of CPMTPg: step1. This code creates a gene pool by step1 of CPMTPg. $\boldsymbol{X}^{(t_r)} = \left(\boldsymbol{x}_1^{(t_r)}, \dots, \boldsymbol{x}_l^{(t_r)}\right)$; $\boldsymbol{x}_j^{(t_r)} = \left(x_j^{(1,t_r)}, \dots, x_j^{(N,t_r)}\right)^{\mathsf{T}}$ $(j = 1, \dots, p)$: gene expression levels of p genes × n subjects at time point t_r. $y^{(i)}$ $(i = 1, \dots, n)$: the therapeutic response of the i^{th} patient. $\lambda^{(k)}$ $(k = 1, \dots, K)$: the k^{th} values of lambda in elastic net in stability selection. $\boldsymbol{G}$ is the gene pool having genes selected by step1 of CPMTPg.

Algorithm:	CPMTPg: Selecting a gene subset
Input:	$\boldsymbol{X} = \{\boldsymbol{X}^{(t_0)}, \dots, \boldsymbol{X}^{(t_R)}\}, \boldsymbol{y} = \{y^{(1)}, \dots, y^{(n)}\}, \boldsymbol{\lambda} = \{\lambda_1, \dots, \lambda_K\}, \boldsymbol{G} \in \{1, \dots, p\}$
Output:	Gene subset $\boldsymbol{G}_{subset} \in \{1, \dots, p\}$
01:	Create $\boldsymbol{X}_{pool}$ by selecting genes of $\boldsymbol{G}$ from $\boldsymbol{X}$
02:	**for** r = 1,...,R **do**
03:	$\boldsymbol{D}_{pool}^{(t_r)} = \boldsymbol{X}_{pool}^{(t_r)} - \boldsymbol{X}_{pool}^{(t_{r-1})}$
04:	Optimize $\boldsymbol{L}^{(t_0 \sim t_r)} \in \mathbf{G}$ using $\{\boldsymbol{D}_{pool}^{(t_1)}, \dots, \boldsymbol{D}_{pool}^{(t_r)}\}$ and $\boldsymbol{y}$ (The fitness function: Eq. 6)
05:	**for** i=1,...,n **do**
06:	Create $\boldsymbol{D}_{learn}^{(t_r)}$ and $\boldsymbol{y}_{learn}$ from $\boldsymbol{D}_{pool}^{(t_r)}$, $\boldsymbol{L}^{(t_0 \sim t_r)}$ and $\boldsymbol{y}$ by selecting n − 1 patients except the i^{th} patient
07:	Create $\boldsymbol{D}_{test}^{(t_r)}$ and $y^{(i)}$ from $\boldsymbol{D}_{pool}^{(t_r)}$, $\boldsymbol{L}^{(t_0 \sim t_r)}$ and $\boldsymbol{y}$ by selecting the i^{th} patient
08:	Optimize $\boldsymbol{w}_{learn}^{(t_r)}$ using $\boldsymbol{D}_{learn}^{(t_r)}$, $\boldsymbol{y}_{learn}$, and $\boldsymbol{\lambda}$ (The elastic Net: Eq. 5)
09:	Estimate $\hat{y}^{(i)}$ using $\boldsymbol{D}_{test}^{(t_r)}$ and $\boldsymbol{w}_{learn}^{(t_r)}$
10:	Calculate $\sigma^{(i)} = \{1 \; if \; \hat{y}^{(i)} = y^{(i)};\; 0 \; if \; \hat{y}^{(i)} \neq y^{(i)}\}$
11:	**for end**
12:	Calculate $acc^{(t_0 \sim t_r)} = (1/N) \sum_{i=1}^{N} \sigma^{(i)}$
13:	**for end**
14:	Select $\boldsymbol{G}_{subset}$ from $\{L^{(t_0 \sim t_1)}, \dots, L^{(t_0 \sim t_R)}\}$ using t_r having the maximum value of $\{acc^{(t_0 \sim t_1)}, \dots, acc^{(t_0 \sim t_R)}\}$

Figure 18 The pseudo-code of CPMTPg: step2. This code creates a gene subset by step2 of CPMTPg. $\boldsymbol{X}^{(t_r)} = \left(\boldsymbol{x}_1^{(t_r)}, \dots, \boldsymbol{x}_l^{(t_r)}\right)$; $\boldsymbol{x}_j^{(t_r)} = \left(x_j^{(1,t_r)}, \dots, x_j^{(N,t_r)}\right)^{\top}$ $(j = 1, \dots, p)$: gene expression levels of p genes × N subjects at time point t_r. $y^{(i)}$ $(i = 1, \dots, N)$: the therapeutic response of the i^{th} patient. $\lambda^{(k)}$ $(k = 1, \dots, K)$: the k^{th} values of lambda in elastic net in stability selection. $\boldsymbol{G}$ is the gene pool having genes selected by step1 of CPMTPg. $\boldsymbol{G}_{subset}$ is the gene subset of CPMTPp.

3.3. Results of CPMTP

The three-fold cross-validation was performed for each HCV and MS dataset. In MLR+maSigPro and CPMTPp+CPMTPg, AUCs, as well as mean maximum and minimum values of accuracies, were calculated based on the results of cross-validation. The mean, maximum, and minimum values of accuracies for MLR+CPMTPg were calculated. Moreover, genes selected via the CPMTPg were analyzed.

The ROC curves and AUCs of MLR+maSigPro and CPMTPp+CPMTPg generated using the HCV dataset are shown (Figure 19). Accordingly, the AUCs of MLR+maSigPro were 0.71 ($CP1$), 0.75 ($CP2$), 0.76 ($CP3$), 0.75 ($CP4$), and 0.76 ($CP5$), respectively. The AUCs of CPMTPp+CPMTPg were 0.89 ($CP1$), 0.90 ($CP2$), 0.90 ($CP3$), 0.90 ($CP4$), and 0.90 ($CP5$), respectively. The p-values of the DeLong test were 0.03 ($CP1$), 0.06 ($CP2$), 0.07 ($CP3$), 0.05 ($CP4$), and 0.06 ($CP5$), respectively. The AUCs of CPMTPp+CPMTPg at all CPs were higher than the AUCs of MLR+maSigPro, and several time points showed a significant difference between these AUC values.

ROC curves and AUCs of MLR+maSigPro and CPMTPp+CPMTPg generated using the MS dataset are shown (Figure 20). The AUCs of MLR+maSigPro from $CP1$ to $CP4$ were 0.76, 0.78, 0.79, and 0.79, while those of CPMTPp+CPMTPg were 0.94, 0.85, 0.91, and 0.93, respectively. The p-values of the DeLong test from $CP1$ to $CP4$ were 0.14, 0.68, 0.38, and 0.30. All AUCs of CPMTPp+CPMTPg were not significantly higher than those of MLR+maSigPro.

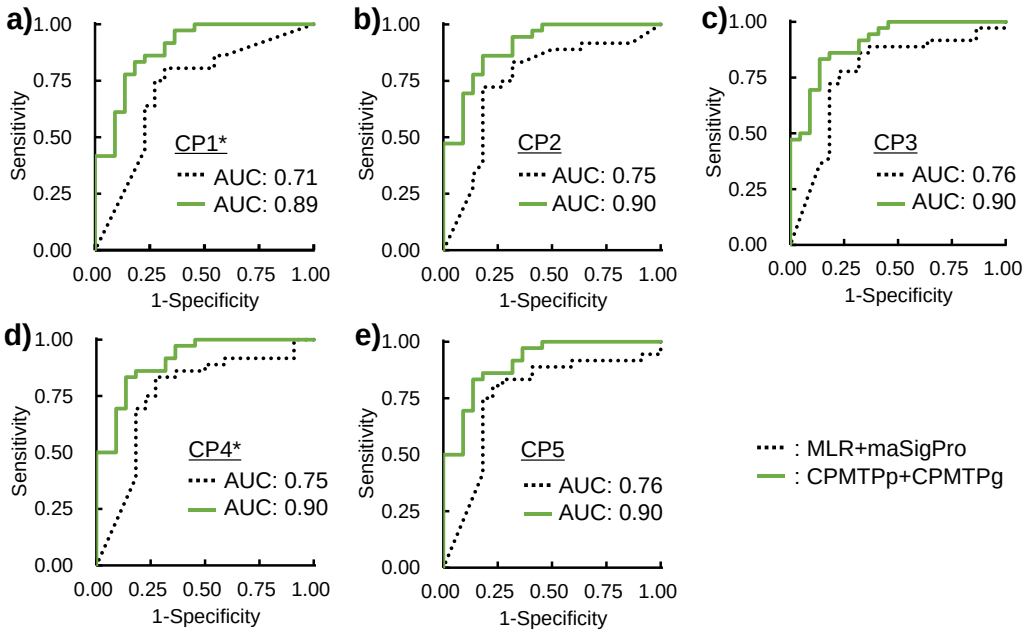

Figure 19 ROC curves of MLR+maSigPro vs. CPMTPp+CPMTPg in HCV dataset. The CP of MLR+maSigPro presents two time points from $\boldsymbol{t_{r-1}}$ to $\boldsymbol{t_r}$. The CP of CPMTPp+CPMTPg presents multiple time points from $\boldsymbol{t_0}$ to $\boldsymbol{t_r}$. "*" means that the difference of AUCs between MLR+maSigPro and CPMTPp+CPMTPg is significant. (a) The case of $CP1$. (b) The case of $CP2$. (c) The case of $CP3$. (d) The case of $CP4$. (e) The case of $CP5$.

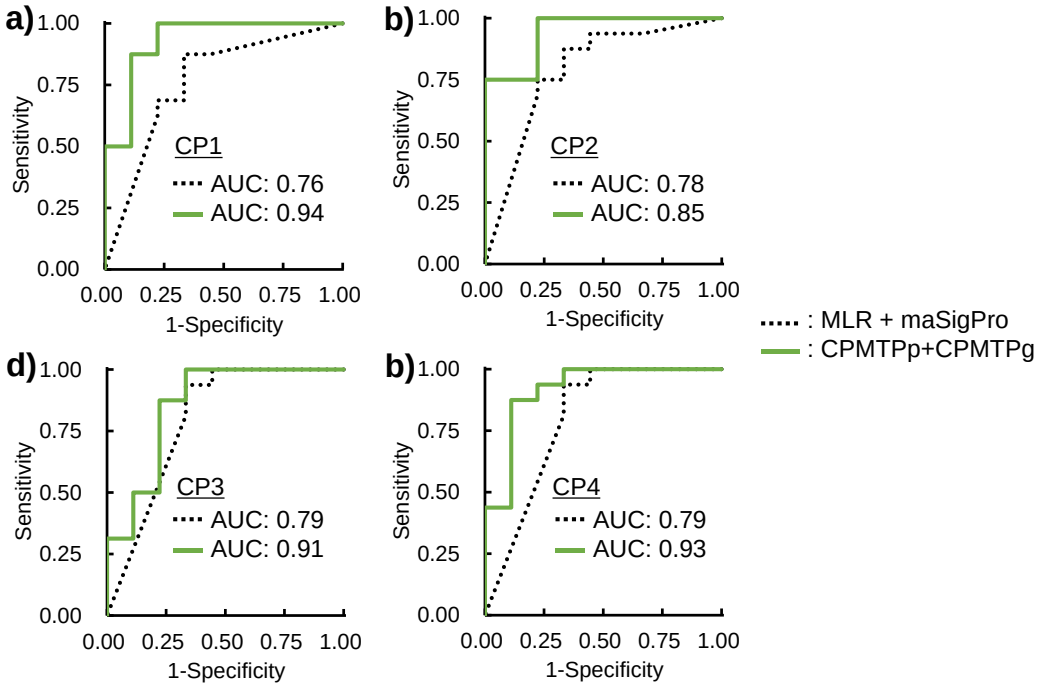

Figure 20 ROC curves of MLR+maSigPro vs. CPMTPp+CPMTPg in MS dataset. The CP of MLR+maSigPro presents two time points from $\boldsymbol{t_{r-1}}$ to $\boldsymbol{t_r}$. The CP of CPMTPp+CPMTPg presents multiple time points from $\boldsymbol{t_0}$ to $\boldsymbol{t_r}$. "*" means that the difference of AUCs between MLR+maSigPro and CPMTPp+CPMTPg is significant. (a) The case of $CP1$. (b) The case of $CP2$. (c) The case of $CP3$. (d) The case of $CP4$.

The accuracies calculated by MLR+maSigPro and CPMTPp+CPMTPg using the HCV dataset are shown (Figure 21 a). The mean accuracies of MLR+maSigPro from $CP1$ to $CP5$ were 74.4%, 63.7%, 61.9%, 62.0%, and 72.5%, respectively. The mean accuracies of the CPMTPp+CPMTPg were 82.8%, 82.8%, 82.8%, 82.8%, and 82.8%, respectively. The minimum and maximum values for the accuracies of MLR+maSigPro from $CP1$ to $CP5$ were 55.0% and 89.4%, 63.1% and 65.0%, 52.6% and 70.0%, 57.8% and 65.0%, and 65.0%, and 84.2%, respectively. The minimum and maximum accuracies of CPMTPp+CPMTPg from $CP1$ to $CP5$ were 80.0% and 84.2%, 75.0% and 89.4%, 75.0% and 89.4%, 75.0% and 89.4%, and 75.0% and 89.4%, respectively. The mean values of CPMTPp+CPMTPg were higher than those of MLR+maSigPro for all CPs. The maximum values for CPMTPp+CPMTPg, with the exception of $CP1$, were higher than those for MLR+maSigPro, while the minimum values at CPs were higher than those for MLR+maSigPro.

The accuracies of MLR+maSigPro and CPMTPp+CPMTPg for the MS dataset are shown (Figure 21 b). The mean accuracies of MLR+maSigPro from $CP1$ to $CP4$ were 75.4%, 68.5%, 71.7%, and 83.3%, respectively. The mean accuracies of CPMTPp+CPMTPg were 83.7%, 83.7%, 87.9%, and 87.9%, respectively. The minimum and maximum accuracies of MLR+maSigPro from $CP1$ to $CP4$ were 62.5% and 88.8%, 55.5% and 75.0%, 62.5% and 77.7%, and 75.0%, and 100.0%, respectively. The minimum and maximum accuracies of the CPMTPp+CPMTPg were 75.0% and 88.8%, 75.0% and 88.8%, 87.5%, and 88.8%, and 87.5% and 88.8%, respectively. The mean values of CPMTPp+CPMTPg were higher than those of MLR+maSigPro for all CPs. The

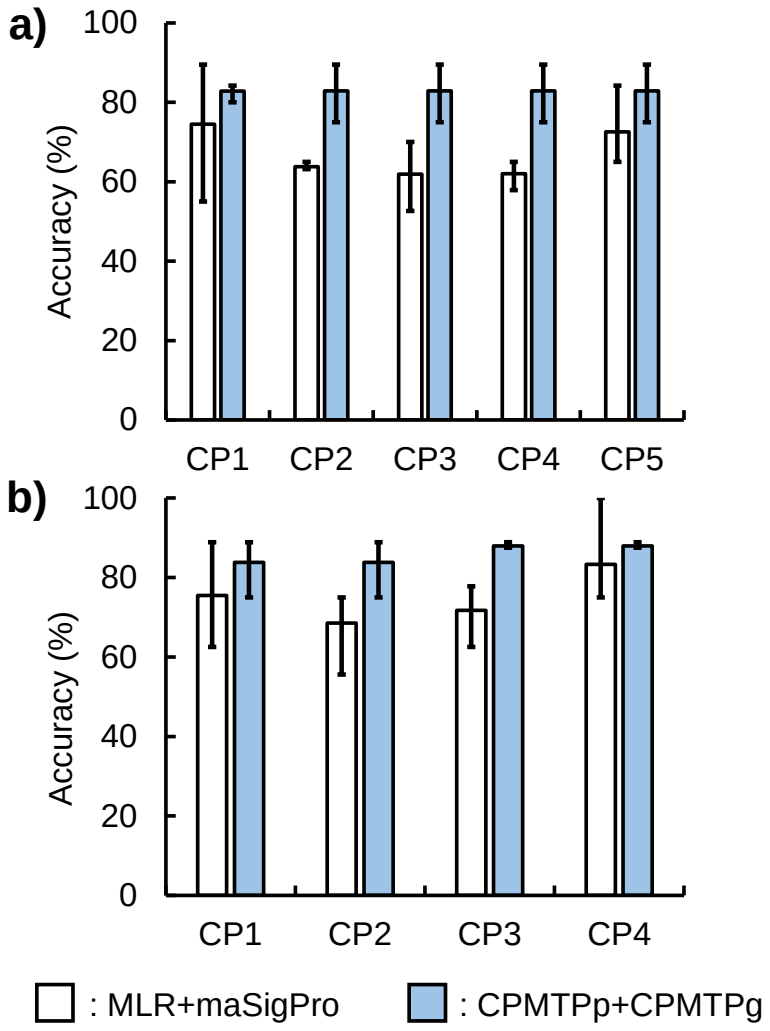

Figure 21 Accuracies of MLR+maSigPro vs. CPMTPp+CPMTPg. The bars, top whisker, and bottom whisker are mean, maximum, and minimum values of accuracies by the three-fold cross-validation, respectively. (a) HCV dataset. (b) MS dataset.

CPs with maximum values for CPMTPp+CPMTPg that were higher than those of MLR+maSigPro were $CP2$ and $CP3$; however, the minimum values of CPMTPp+CPMTPg at all CPs were higher than those of MLR+maSigPro.

The accuracies of MLR and CPMTPp estimated using the gene subset selected from the HCV dataset via CPMTPg are shown (Figure 22 a). The mean values of accuracies estimated by MLR+CPMTPg were 82.8 ($CP1$), 60.3 ($CP2$), 63.8 ($CP3$), 62.1 ($CP4$), and 62.1 ($CP5$), respectively. The minimum and maximum values of accuracies estimated by MLR+ CPMTPg were 80.0% and 84.4% ($CP1$), 57.8% and 63.1% ($CP2$), 60.0% and 68.4% ($CP3$), 60.0% and 63.1% ($CP4$), and 60.0% and 63.1% ($CP5$), respectively. The mean, maximum, and minimum values of accuracies estimated by CPMTPp+ CPMTPg were the same as those shown in Figure 22 a. The accuracy of MLR+CPMTPg at $CP1$ was highest, while the accuracies for the other CPs decreased. On the other hand, the accuracy of CPMTPp+CPMTPg did not change with the increase in CPs.

MLR and CPMTPp were compared for accuracy using the MS subset (Figure 22 b). The gene subsets of MLR and CPMTPp were common. The mean values of the accuracies of MLR+CPMTPg were 83.7 ($CP1$), 72.2 ($CP2$), 84.2 ($CP3$), and 67.5 ($CP4$), respectively. The minimum and maximum values of accuracies of MLR+ CPMTPg were 75.0% and 88.8% ($CP1$), 66.6% and 75.0% ($CP2$), 75.5% and 100.0% ($CP3$), and 62.5% and 77.7% ($CP4$), respectively. The mean, maximum, and minimum values of accuracies of CPMTPp+CPMTPg were the same as those shown in Figure 22 b. The accuracy of MLR+ CPMTPg at $CP1$ were different at each CP. The accuracy of CPMTPp+ CPMTPg slightly improved as CPs increased.

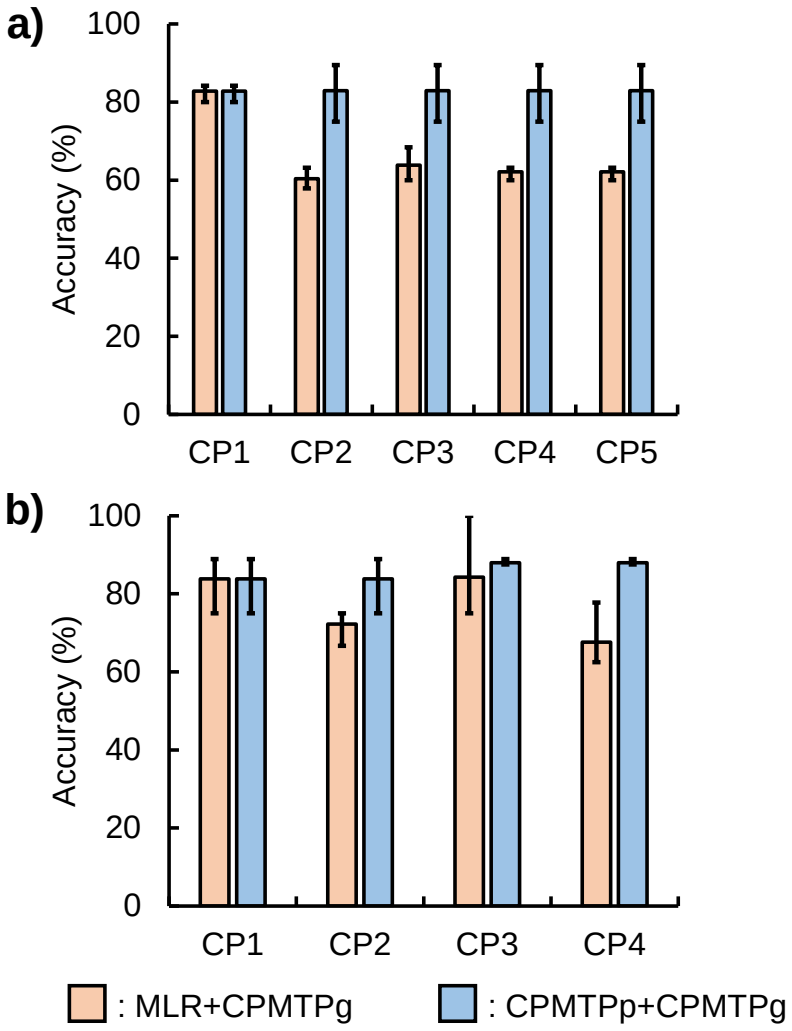

Figure 22 Accuracies of MLR+CPMTPg vs. CPMTPp+CPMTPg. The bars, top whisker, and bottom whisker are mean, maximum, and minimum values of accuracies by the three-fold cross-validation, respectively. (a) HCV dataset. (b) MS dataset.

The mean accuracies of CPMTPp+CPMTPg using the HCV dataset were not changed as time progressed (Figure 22 a). However, the mean accuracies of the MS dataset improved slightly with increasing time (Figure 22 b). Further, the maximum and

minimum values either did not change or improved slightly. Thus, in contrast to our hypothesis, the accuracies estimated using the two datasets either did not change or improved slightly with increasing time.

In the HCV dataset, 30 genes were selected by CPMTPg as the gene subset for the logistic regression model from the learning data on the three-fold cross-validation. The GO terms of the HCV dataset, which were determined by these genes, generated 4 clusters. The 10 GO terms had significant p-values (Table 11). “Repeat: 1”, “Repeat: 2”, and “Repeat: 3”, which belonged to the same cluster and were selected by the same genes, were not terms associated with gene function. “Proteinaceous extracellular matrix,” “Disulfide bond,” and “Extracellular matrix” belonged to the same cluster, which was not the top cluster. “Disease mutation,” “Polymorphism,” “Visual perception,” and “Positive regulation of transcription, DNA-templated” did not belong to any cluster.

Twenty-six genes were selected by CPMTPg using the MS dataset, where 4 were selected twice in the three-fold cross-validation. The GO terms of the MS dataset were decided according to these genes, and 3 clusters were constructed (Table 12). The GO terms with significant p-values are shown. “Fatty acid metabolism” belonged to the cluster, while “Nucleus” and “protein binding” did not belong to any cluster.

Table 11 Selected GO terms in the HCV dataset.

GO term	Genes	Count	p-value	Cluster
Repeat:3	*ADAM30, GFRA1, PSRC1*	3	0.036	#1
Repeat:1	*ADAM30, GFRA1, PSRC1*	3	0.046	#1
Repeat:2	*ADAM30, GFRA1, PSRC1*	3	0.047	#1
Proteinaceous extracellular matrix	*EFEMP1, WNT5A, KERA, OLFML2B*	4	0.007	#2
Disulfide bond	*ADAM3, EFEMP1, GFRA1, KLRC4-KLRK1, WNT5A, CACNA1A, IGLL1, KERA, OLFML2B, PRPH2*	10	0.026	#2
Extracellular matrix	*EFEMP1, WNT5A, KERA*	3	0.045	#2
Disease mutation	*EFEMP1, WNT5A, ACAT1, CACNA1A, CCND2, IGLL1, KERA, PRPH2, KCNK3, SRD5A2*	10	0.039	Not belong
Polymorphism	*AKAP5, ADAM30, EFEMP1, GFRA1, KLRC4-KLRK1, MAGEA10, ACAT1, CACNA1A, CAMTA1,*	22	0.030	Not belong

	CCND2, EIF3F, IGLL1, MED24, OLFML2B, OGDHL, PRPH2, PSRC1, PCDHGA3, SRD5A2, ZNF43, ZNF512B, ZNF711			
Visual perception	*EFEMP1, KERA, PRPH2*	3	0.036	Not belong
Positive Regulation of transcription, DNA-templated	*WNT5A, MED24, PSRC1, ZNF711*	4	0.040	Not belong

These terms have lower p-values than 0.05 (significance level). Thirty-one GO terms belong to four clusters. On the other hand, 13 GO terms do not belong to any cluster. The clusters were generated during the GO analysis.

Table 12 Selected GO terms for the MS dataset.

GO term	Genes	Count	p-value	Cluster
Fatty acid metabolism	*NDUFAB, ACAA2, ALOX15*	3	0.008	#1
Nucleus	*LARP6, RBM47, CENPO, ESRRA, MTDH, MORF4L1, PA2G4, RSL24D1, ZBED1, ZNF516, ZNF614*	11	0.033	Not belong
Protein binding	*NDUFAB1, ACAA2, ALOX15, CENPO, ESRRA, MTDH, MAT2A, MORF4L1, PA2G4, RSL24D1, SERPINA5, TRPC3, TRPM8, ZBED1, ZNF614*	15	0.047	Not belong

These terms have lower p-values than 0.05 (significance level). Twenty-eight GO terms belong to three clusters. On the other hand, 21 GO terms do not belong to any cluster. The clusters were generated during the GO analysis.

3.4. Discussion of CPMTP

AUCs and accuracies calculated using our proposed method (CPMTPp+CPMTPg) were compared with those calculated using the conventional method (MLR+maSigPro) via the three-fold cross-validation. The results of both AUCs (Figure 19 and Figure 20) and accuracies (Figure 21) suggested that our method could predict response to therapy accurately at multiple time points compared to the conventional method.

The AUCs of CPMTPp+CPMTPg were higher than those of MLR+ maSigPro for all CPs in both the HCV and MS datasets (Figure 19 and Figure 20). However, CPs that showed significant differences were $CP1$ and $CP4$ in the HCV dataset, while the differences in the MS dataset were not significant for any CP. This is due to the insufficient number of patients to perform the DeLong test, especially in the MS dataset, where the patient number was 25. Almost all CPs did not show a significant difference; however, a common trend in both HCV and MS datasets was that the AUCs of CPMTPp+CPMTPg were higher than those of MLR+ maSigPro at all CPs.

According to Figure 21, the mean accuracies of CPMTPp+CPMTPg were higher than those of MLR+maSigPro at all CPs, an observation common to both datasets. Moreover, the mean accuracies of CPMTPp+CPMTPg at each CPs were higher than the 72.4% cited in the reference [58] using the same HCV dataset. In the MS dataset, the mean accuracies of CPMTPp+CPMTPg at all CPs were also higher than the 78.0% cited in the reference [34].

In addition, the accuracies of CPMTPp+CPMTPg were confirmed for the artificial data (Figure 23). The results are shown in. The mean accuracies were more than 90.0% at all CPs.

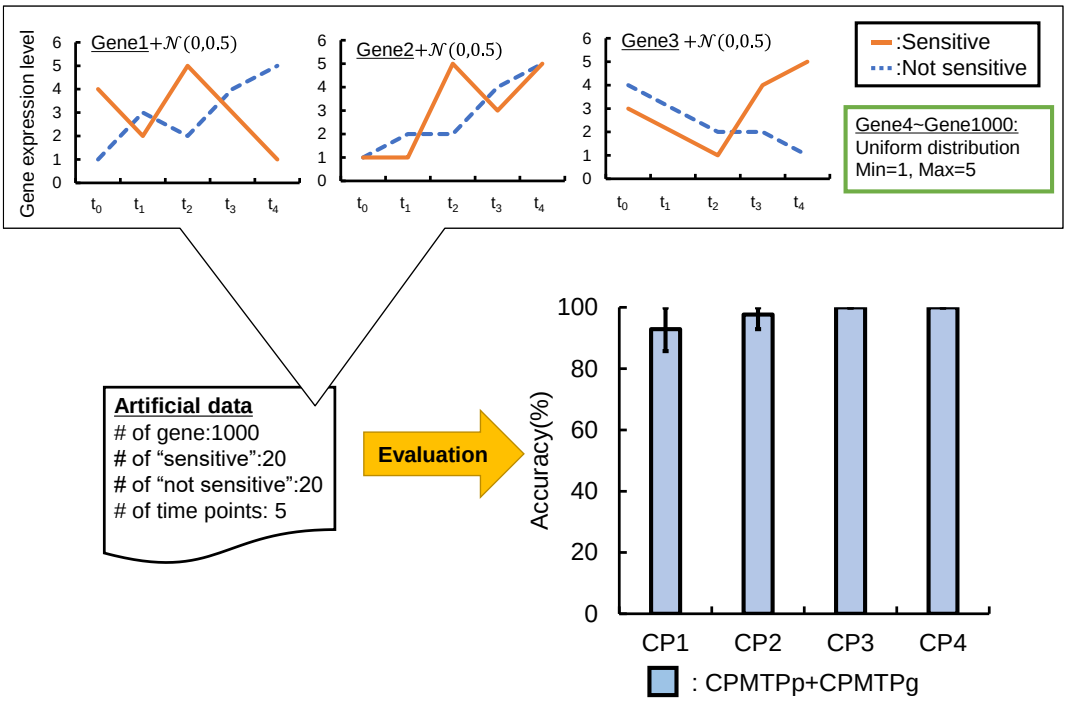

Figure 23 Results of CPMTPp+CPMTPg using artificial data. The artificial gene expression data (1000 genes × 40 subjects × 5 time points; "#" in this figure means "number") were created. These data subjects are 20 sensitive and 20 not sensitive responders. Gene expression levels of *Gene1*, *Gene2*, and *Gene3* were created by adding noise following a normal distribution (center:0; standard deviation:0.5) to each baseline. The baseline of *Gene1* has the different rising/ falling trends of gene expression levels between sensitive and not sensitive responders at all time points, while the baseline of *Gene2* and *Gene3* had it at a part of time points. Gene expression levels of the other genes were created by uniform distribution (maximum:1; minimum:5). To evaluate CPMTPp+CPMTPg, the three-fold cross-validation was performed using the artificial data. As a result, CPMTPg

these mean accuracies were 92.8% ($CP1{:}t_0{\sim}t_1$), 97.6% ($CP2{:}t_0{\sim}t_2$), 100% ($CP3{:}t_0{\sim}t_3$), and 100% ($CP4{:}t_0{\sim}t_4$), respectively. The accuracy at the early term was higher than 90%, and this value increased along with the time progressing. Similar trends were observed using actual datasets in this paper.

CPMTPp was designed based on the hypothesis that more accurate prediction was dependent on data from more time points. However, the results of the comparison between MLR and CPMTPp (Figure 22) did not support this hypothesis, although it indicated that CPMTPp continued to maintain accuracies as time points increased.

The accuracies of MLR, which did not consolidate the probabilities at multiple time points in the HCV and MS datasets, are shown (Figure 22). In the HCV dataset, the top CP, which corresponded to the highest mean accuracy of MLR+CPMTPg, was $CP1$, after which the mean values corresponding to $CP2{\sim}CP5$ decreased (Figure 22 a). The mean accuracies of MLR+CPMTPg for various CPs of the MS dataset appeared to be uncorrelated (Figure 22 b). The trends were also different regarding the maximum and minimum values. When the probabilities at multiple time points were not used for prediction as time points increased, the accuracies did not change or improve as in CPMTPp but were reduced or disjointed.

The above results indicated that prediction using more time points (CPMTPp) did not contribute to improved accuracy. However, MLR, which did not consolidate the probabilities of multiple time points, used the same subset of genes as CPMTPp, and its accuracy tended to decrease or fluctuate overtime points. This trend was not changed by the gene selection method for maSigPro (Figure 24). Therefore, it was found that the

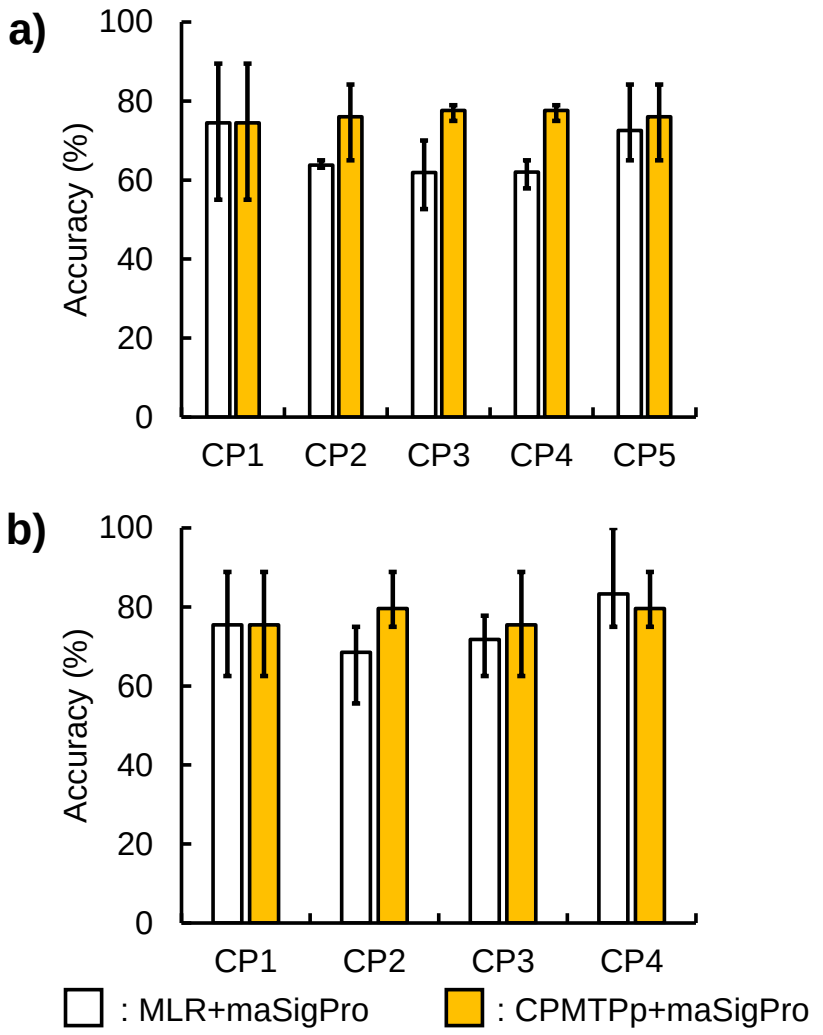

Figure 24 Accuracies of MLR+maSigPro vs. CPMTPp+maSigPro. The bars, top whisker, and bottom whisker presents mean, maximum, and minimum values of accuracies by three-fold cross-validation, respectively. (a) HCV dataset. (b) MS dataset.

accuracies of CPMTPp contributed to maintaining accuracies as time points were processed, in contrast to MLR.

The gene subsets selected by CPMTPg were analyzed, and GO terms were extracted from the DAVID database (Table 11 and Table 12). Genes associated with terms that reportedly played an important role in diseases were discovered by reviewing previous studies that cited significant GO terms.

The GO terms (Table 11) included those that were reportedly associated with HCV infection. The extracellular matrix has been reported to develop progressive hepatic fibrosis and cirrhosis in 20% to 30% of HCV patients [113]. Previous studies have suggested that angiotensin II [113] and fibrogenic cytokines [114] contributed to the production of extracellular matrix in the liver. It was reported that excessive accumulation of extracellular matrix components, such as fibrillar type I and III collagens, fibronectin, and laminin, is a feature of liver fibrosis [115][116]. Another study reported that the accumulation of extracellular matrix in liver fibrosis might impair the signaling of interferon used as therapy [115]. Regarding disulfide bonds, it was reported that a disulfide bond core protein complex might constitute the nucleocapsid-like particle of HCV [117].

The GO terms (Table 12) included those reported to be related to MS. It was suggested that “Fatty acid metabolism” may be a target for MS therapy since inhibition of carnitine palmitoyltransferase 1 (*CPT-1*), which is the rate-limiting enzyme in the beta-oxidation of fatty acids, contributes to a reduction in disease severity [118]. Especially, it was reported that when the *ALOX15* gene, which encodes a fatty acid metabolizing enzyme, became functionally inactive, MS patients experienced more severe symptoms than when the *ALOX15* gene was active [119][120].

The results of these numerical experiments using HCV and MS datasets suggested that CPMTP, our proposed method, may predict responses to therapy more accurately than the conventional method at multiple time points. Besides, CPMTP was able to select genes with functions associated with diseases from time-series microarray data.

CPMTP could be applied to gene expression data with arbitrarily selected multiple time points, and increasing time points did not affect the prediction model of CPMTP. CPMTP could be performed beyond the last time-point of treatment; however, it required validation. Moreover, CPMTP could be applied to RNA-seq data and other gene expression data, which used a normalization similar to log2 fold-change and quantile normalization. When the proposed method is applied to the relatively large database, parameters and optimization methods, such as number of genes, number of samples, and number of time points, should be carefully considered.

In individual patients showing specific therapeutic effects or occasional side effects, it is essential to accurately predict response to therapy using gene markers to determine a therapeutic strategy, such as changing or stopping therapy. Here, we propose a new prediction model and gene selection method termed CPMTP, which comprises a prediction component and selection component. CPMTP was based on the hypothesis that more information related to time points provided a more accurate therapeutic response prediction. To enable CPMTPp to incorporate more information from multiple time points, an overall probability of deciding a therapeutic response was estimated by consolidating the probabilities calculated at each time point, using the Bayesian theorem. CPMTPg selected the gene subset for use in the CPMTPp model via the optimization method, which was set as the fitness function of the consolidated probability.

CPMTP was evaluated using time-course gene expression profiles from HCV and MS patients in terms of accurate prediction, validation of the hypothesis, and gene function. These results suggested that CPMTP predicted response to therapy accurately at all observed points compared to the conventional method. However, as opposed to our hypothesis, the predicted accuracy of CPMTPp was not improved but only retained as time points increased. Further, the gene subset selected by CPMTPg may be related to HCV and MS, according to analyses conducted by previous studies investigating the key GO terms associated with the gene subsets.

The above findings indicated that CPMTP might enhance long-term therapeutic procedures by accurately predicting response to therapy at multiple time points. Moreover, gene subsets identified by CPMTP may be useful as gene markers of disease progression. Thus, CPMTP may not only resolve difficulties associated with predicting response to therapy in HCV and MS patients but may also apply to the resolution of other clinical issues of a similar nature.

4. Comparison of eENCD and CPMTP

The present study introduced two new methods, eENCD and CPMTP, for biomarker discovery. Both methods challenged the problems of (1) the $p \gg n$ problem, (2) multicollinearity, and (3) accurate prediction at multiple time points. These methods selected genes with different patterns of time-course gene expression profiles. The eENCD selected genes showing consistent differentiation between the two groups at all time points. However, CPMTP was designed to select genes that always showed a difference in gene expression level between the two groups, and predicted the therapeutic response using the predicted results of multiple time points. In this section, eENCD and CPMTP are compared from the point of view of predictive performance and gene function.

In the first evaluation, the predicted accuracies and AUCs for the predictive performance of eENCD and CPMTP were compared using the GEO dataset ID of GSE24427. The predicted accuracies of eENCD shown in Figure 12 were 70.4% ($CP1$: t_1), 82.8% ($CP2$: t_2), 80.8% ($CP3$: t_3), and 78.3% ($CP4$: t_4). As shown in Figure 21, the predicted accuracies of CPMTP were 83% ($CP1$), 83% ($CP2$), 87% ($CP3$), and 87% ($CP4$). The predicted accuracies at all time points from the CPMTP method were higher than those predicted using the eENCD method. In addition, eENCD maintained the predicted accuracies at all time points, whereas the CPMTP accuracies improved a bit with an increasing number of time points.

The AUCs of eENCD and CPMTP were compared using the same dataset to evaluate the predicted accuracies. As shown in Figure 7 b, the AUCs of eENCD were 0.76 ($CP1$: t_1), 0.95 ($CP2$: t_2), 0.89 ($CP3$: t_3), and 0.95 ($CP4$: t_4). The AUCs of CPMTP from Figure 20 were 0.94 ($CP1$), 0.85 ($CP2$), 0.91 ($CP3$), and 0.93 ($CP4$). The AUCs of

CPMTP were equal to and lower than those of eENCD for almost all CPs. As the number of time points increased, the AUCs of eENCD varied regardless of CP, whereas those of CPMTP increased at multiple CPs, except for $CP2$.

Therefore, the predicted accuracies and AUCs of CPMTP were higher than those of eENCD for almost all time points. As the number of time points increased, the prediction accuracies and AUCs of CPMTP improved a bit. Therefore, the predictive performance of CPMTP was greater than that of eENCD.

In the second evaluation, the genes selected using the CPMTP and eENCD methods were compared based on the gene expression level patterns and functions, using the same dataset as the first evaluation. As shown in Figure 10 b, the gene expression levels of the *CDH2* gene selected with the eENCD method showed consistent differentiation between sensitive and non-sensitive responders for all time points. The gene expression levels of the *ACAA2*, *PA2G4, SERPINA5,* and *MORF4L1* genes selected by CPMTP are shown in Figure 25, Figure 26, Figure 28, and Figure 27, respectively. These genes showed that different gene expression levels at the early time points differed significantly between sensitive and non-sensitive responders. These results did not support the hypothesis of CPMTP that more accurate prediction requires more time points and revealed that gene expression levels at an early time point strongly affected the therapeutic response decision. Thus, the collection of gene expression profiles at early time points is very important for the predictive performance of prognostic and predictive biomarkers.

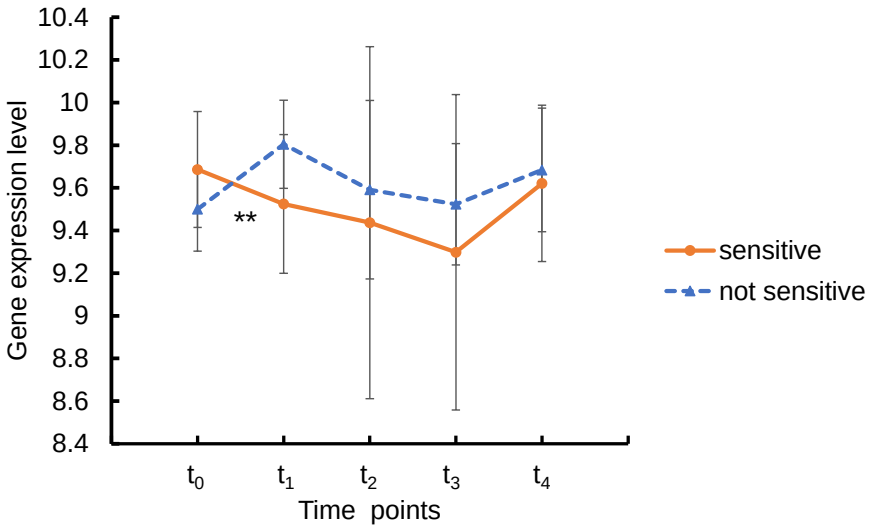

Figure 25 Gene expression profiles of the *ACAA2* in the MS dataset.

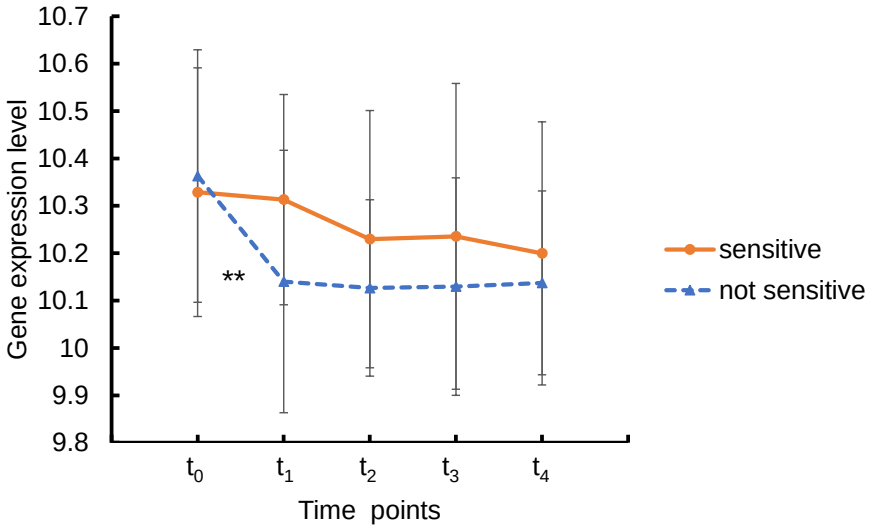

Figure 26 Gene expression profiles of the PA2G4 in the MS dataset. This gene was selected by CPMTP twice in the three-fold cross-validation. The p-values at each CP were calculated with the Wilcoxon sum-rank test. The "**" means that the p-value was less than 0.01. The "*" means that the p-value was less than 0.05.

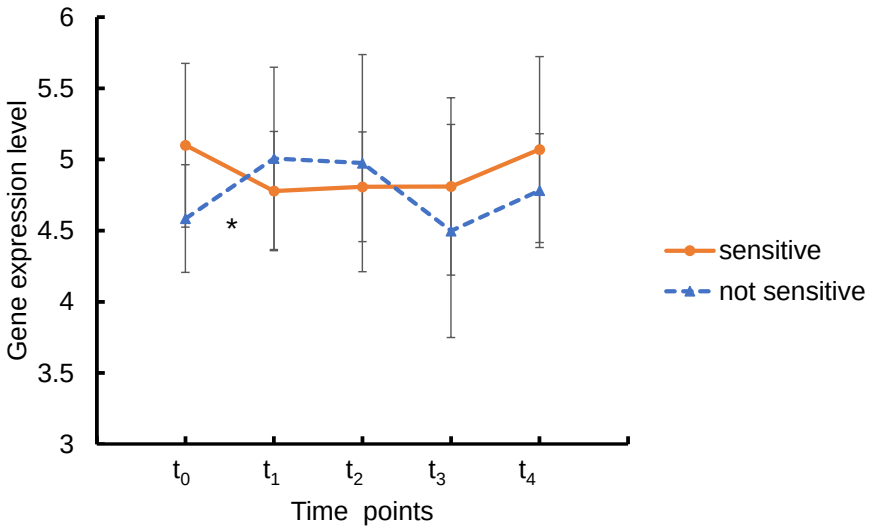

Figure 27 Gene expression profiles of the *SERPINA5* in the MS dataset.

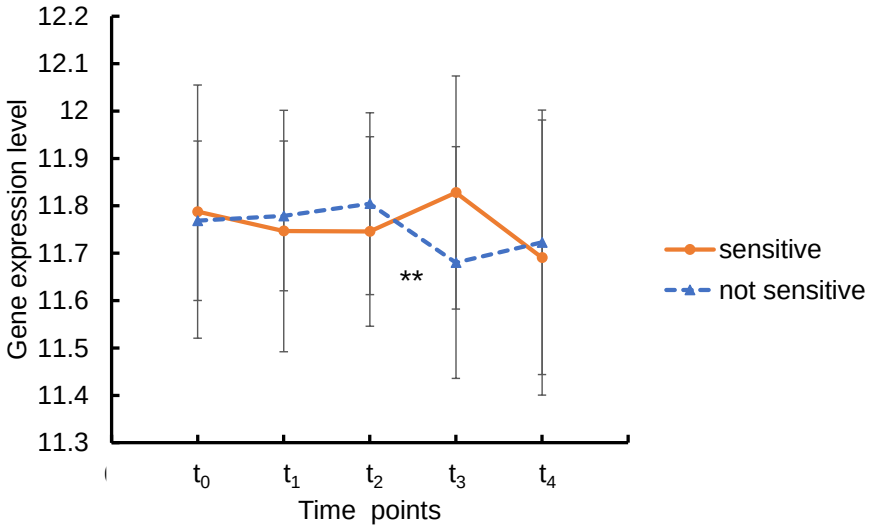

Figure 28 Gene expression profiles of the *MORF4L*1 in the MS dataset.

This gene was selected by CPMTP twice in the three-fold cross-validation. The p-values at each CP were calculated with the Wilcoxon sum-rank test. The "**" means that the p-value was less than 0.01. The "*" means that the p-value was less than 0.05.

Table 13 GO terms in the cluster with a significant enrichment score using genes selected by eENCD and CPMTP.

GO term	Genes of eENCD	Genes of CPMTP	p-value
Fatty acid metabolism	*FLAD1*	*NDUFAB1, ACAA2, ALOX15*	0.009
Metabolic pathways		*NDUFAB1, ACAA2, ALOX15, MAT2A*	0.059
Lipid metabolism		*NDUFAB1, ACAA2, ALOX15*	0.084

The 27 genes selected by CPMTP and eENCD in Table 6 and Table 12 were analyzed with gene ontology (GO) using the DAVID tool. Genes with similar GO terms were classified into the same cluster, and the number of clusters was six. In these clusters, there was one cluster whose enrichment score was over 1.3, which was significant. The GO terms belonging to this cluster are shown in Table 13. The p-value of fatty acid metabolism was 0.008, which was lower than the 0.05 significance level, whereas the p-values of metabolic pathways and lipid metabolism were not lower than the significance level. However, these GO terms were related to metabolism. In Table 13, *FLAD1*, *NDUFAB1*, *ACAA2*, *ALOX15*, and *MAT2A* genes, which were related to the metabolic pathways of the GO term, were analyzed using KEGG (https://www.genome.jp/kegg/). The *FLAD1* gene was selected with eENCD, and the other genes were selected using CPMTP. In the metabolic pathway whose ID was "hsa01100," the pathways related to *FLAD1*, *NDUFAB1*, *ACAA2*, *ALOX15*, and *MAT2A* genes were "riboflavin metabolism" (Figure 29), "oxidative phosphorylation" (Figure 30), "fatty acid degradation" (Figure

31), “linoleic acid metabolism” (Figure 32), and “cysteine and methionine metabolism” (Figure 33), respectively.

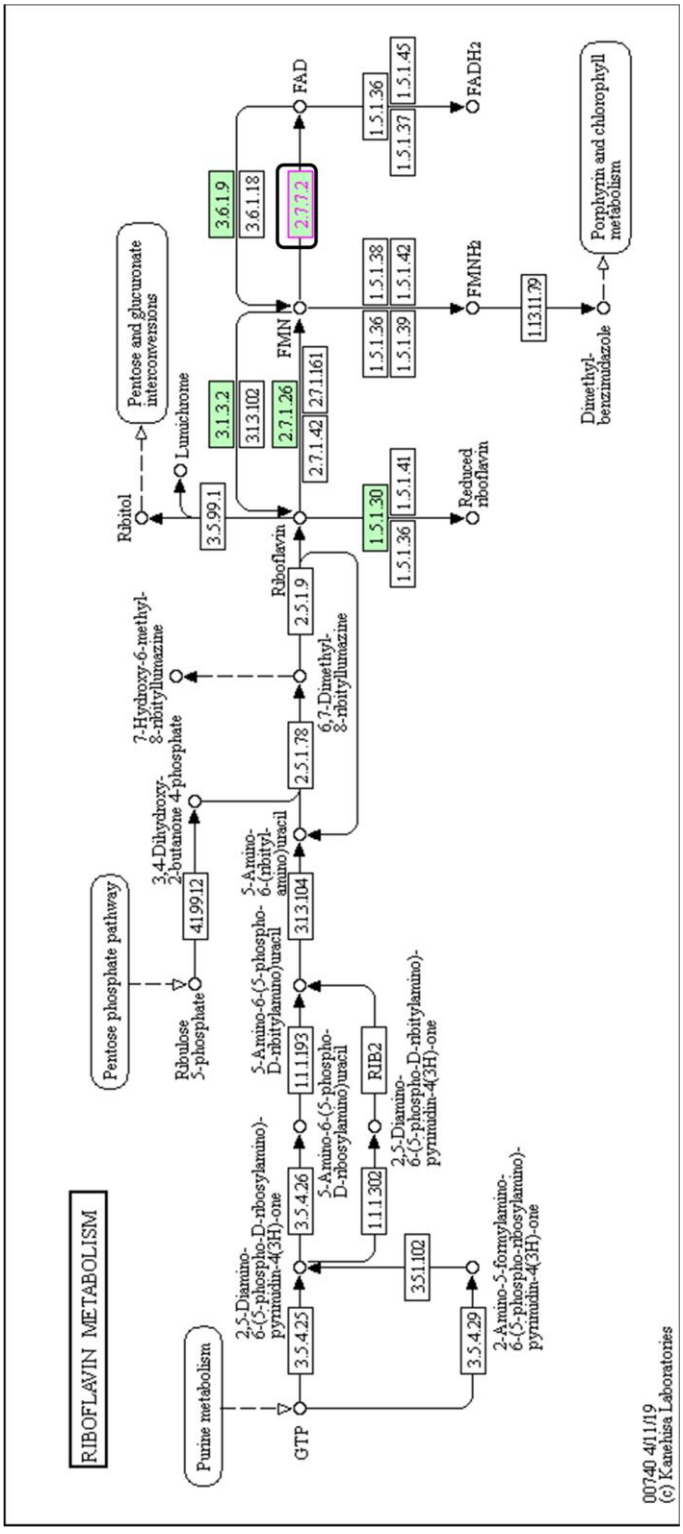

Figure 29 Riboflavin metabolism pathway related to the *FLAD1* (Source: KEGG, https://www.genome.jp/kegg/). The black-colored box means the relationship of the *FLAD1* gene in this pathway.

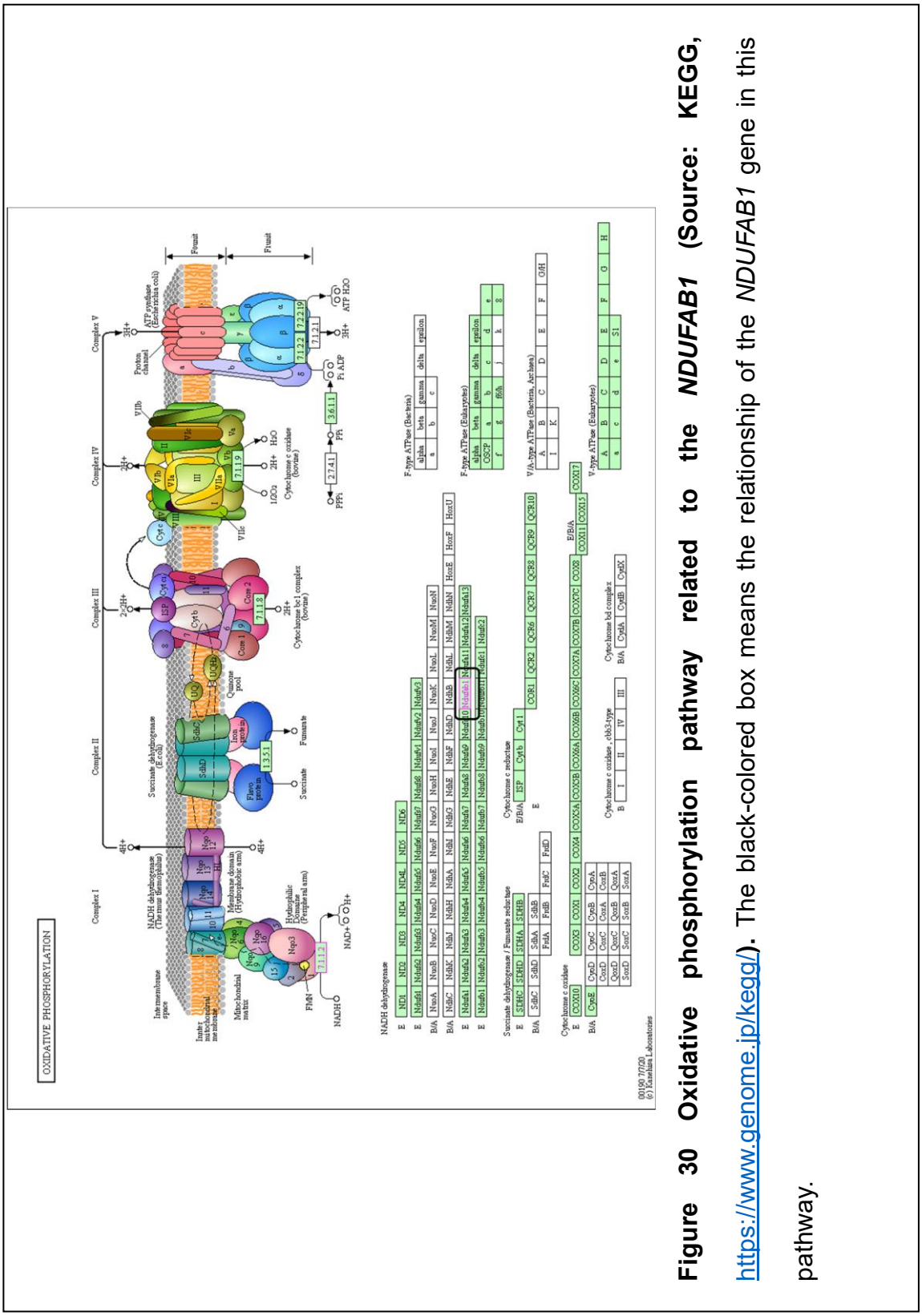

Figure 30 Oxidative phosphorylation pathway related to the *NDUFAB1* (Source: KEGG, https://www.genome.jp/kegg/). The black-colored box means the relationship of the *NDUFAB1* gene in this pathway.

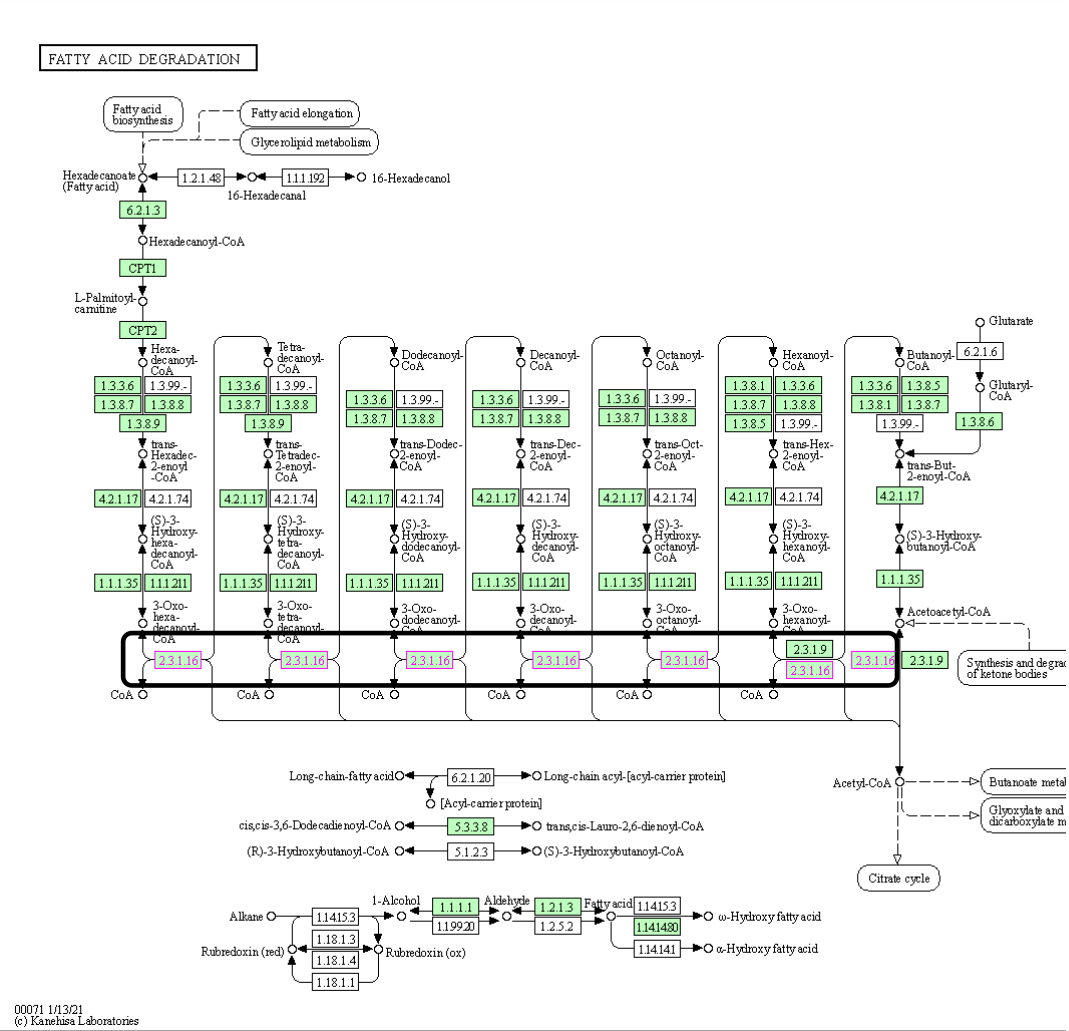

Figure 31 Fatty acid degradation pathway related to the *ACAA2* (Source: KEGG, https://www.genome.jp/kegg/). The black-colored box means the relationship of the *ACAA2* gene in this pathway.

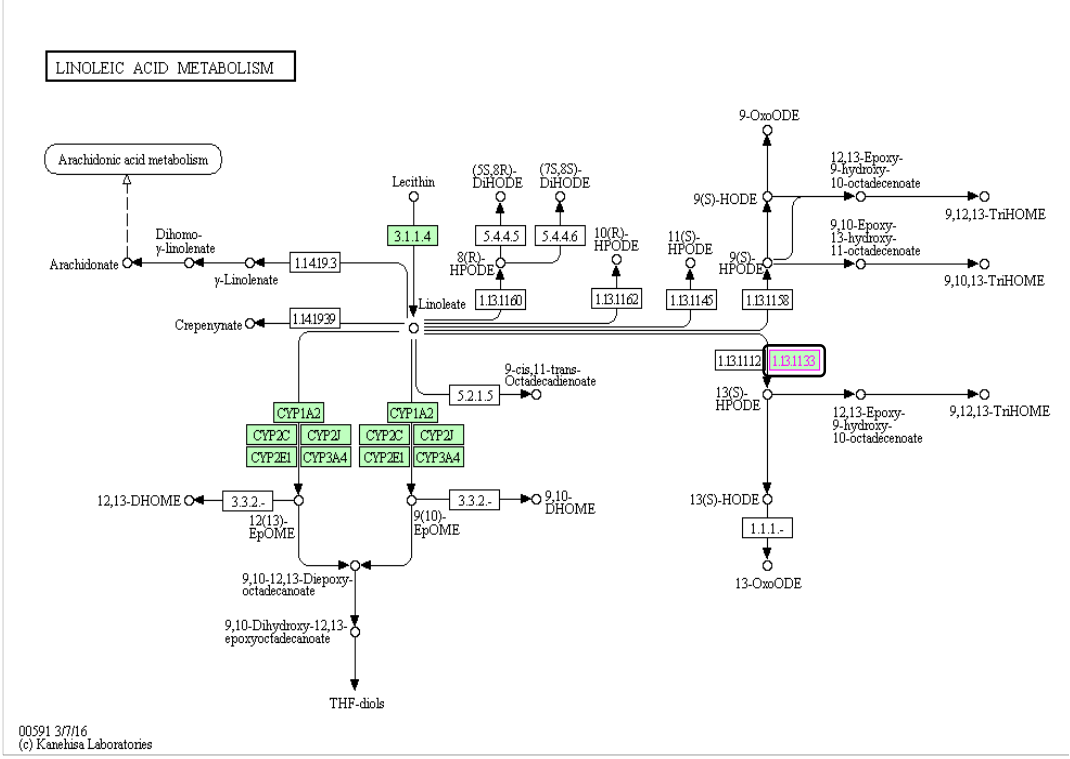

Figure 32 Linoleic acid metabolism pathway related to the *ALOX15* (Source: KEGG, https://www.genome.jp/kegg/). The black-colored box means the relationship of the *ALOX15* gene in this pathway.

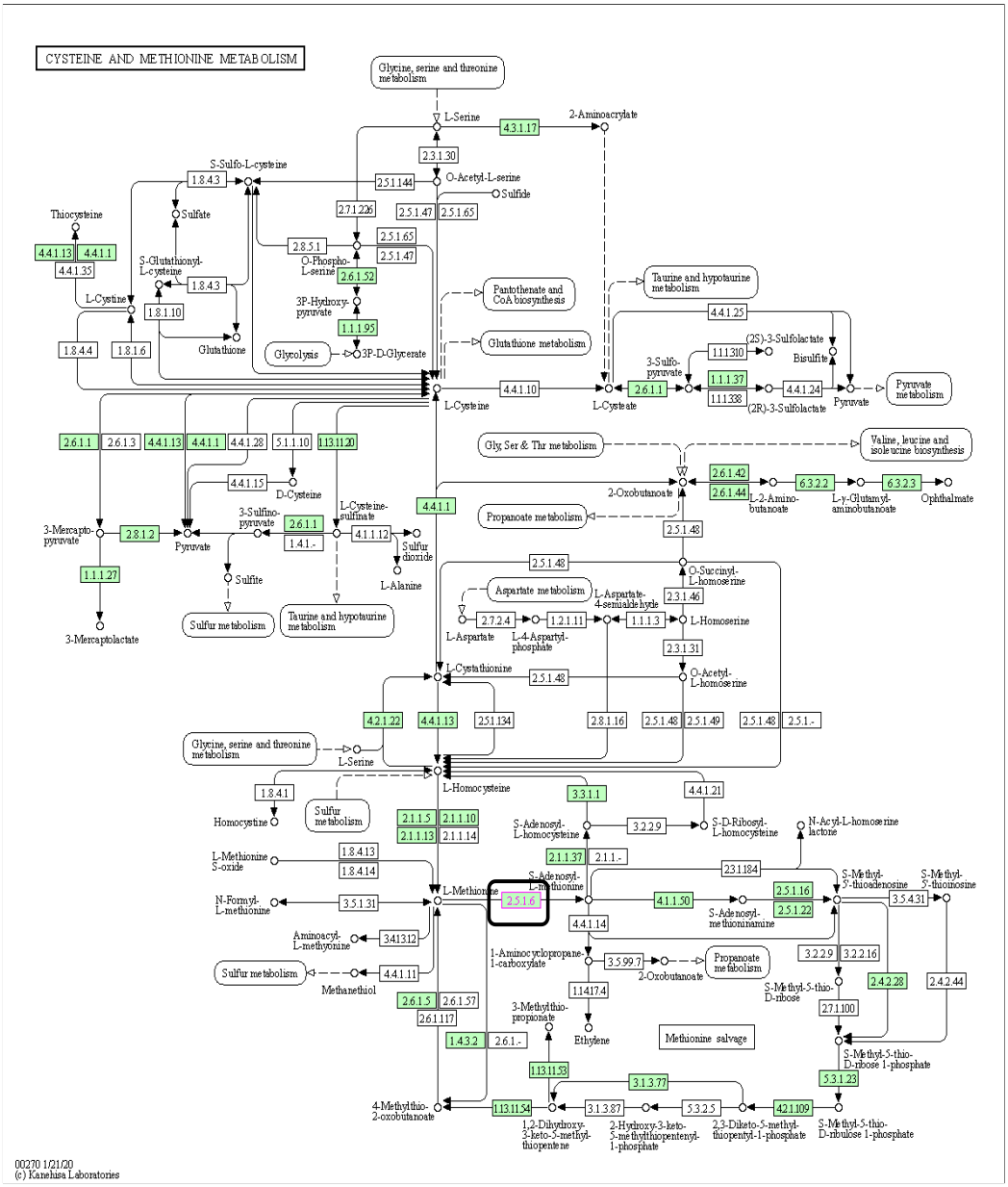

Figure 33 Cysteine and methionine metabolism pathway related to the *MAT2A* (Source: KEGG, https://www.genome.jp/kegg/). The black-colored box means the relationship of the *MAT2A* gene in this pathway.

Riboflavin metabolism, oxidative phosphorylation, fatty acid degradation, linoleic acid metabolism, and cysteine and methionine metabolism were investigated for their association with MS in ID GSE24427. Riboflavin is associated with several neurological disorders due to mitochondrial dysfunction, and MS is one such disorder [121]. In an experiment using rat livers, riboflavin deficiency caused disruption of the myelin lamellae, and the mitochondrial fatty acid oxidation was suppressed and restored after supplying with riboflavin [122]. Because of this deficiency, the dehydrogenation of fatty acids is reduced, resulting in linoleic, linolenic, and arachidonic acid reduction [123]. Oxidative phosphorylation is a reaction catalyzed by multiprotein complexes encoded by the nucleus and mitochondria [124], and improves the decrease in muscle strength and exercise tolerance in patients with defects in oxidative phosphorylation [125][126][127]. Duatte et al. reported that transcripts related to oxidative phosphorylation were altered in the motor cortex of patients with MS [124]. Linoleic acid is significantly lower in erythrocyte ghosts in patients with MS [128]. Low glucose in fatty acid metabolism promotes fatty acid degradation into acetyl-CoA (beta-oxidation), which inhibits axon myelination, leading to the destruction of the central nervous system [129]. Homocysteine is associated with various neurological disorders [130], and Monti et al. reported that the administration of homocysteine had a positive effect on glucose metabolism in the brain of patients with MS [131].

Thus, riboflavin metabolism, oxidative phosphorylation, fatty acid degradation, linoleic acid metabolism, and cysteine and methionine metabolism were related to *FLAD1*, *NDUFAB1*, *ACAA2*, *ALOX15*, and *MAT2A* genes, respectively, associated with the biological process of MS. The number of genes selected with CPMTP, which were related

to MS, was higher than the number of genes selected with eENCD. In particular, fatty acid degradation related to the *ACCA2* gene selected by CPMTP might be a potential therapeutic target for MS [132]. Therefore, CPMTP selected genes that provided more beneficial information about MS than eENCD.

The predictive performance and function of genes selected by eENCD and CPMTP were compared using time-course gene expression profiles of MS. The predictive performance of CPMTP was higher than that of eENCD for almost all time points. CPMTP selected more genes reporting a relationship with MS than eENCD. Therefore, CPMTP might be better than eENCD as a biomarker discovery method. Gene expression levels at early time points are very important for accurate predictive performance because of breaking the hypothesis with CPMTP that more time points used in prediction improved the predictive performance.

5. Conclusion

Time-course gene expression profiles using microarrays are more valuable than those collected at a single time point. Therefore, it is considered that biomarkers using time-course gene expression profiles can predict therapeutic responses more accurately. However, biomarker discovery using time-course gene expression profiles is complex and challenging for three reasons: (1) the $p >> n$ problem, (2) multicollinearity, and (3) accurate prediction at multiple time points. In this study, we challenged these problems and introduced two new methods for biomarker discovery: eENCD and CPMTP.

In eENCD, the elastic net is expanded to select genes showing consistently distinct gene expression levels between the two groups at all time points for the problem (3). The elastic net is a type of sparse modeling method that solves problems (1) and (2). Evaluating eENCD with the two datasets collected from MS patients, the predictive performance of genes selected by eENCD was higher than that selected by the conventional method. These genes are related to MS, according to previous studies. However, eENCD has two problems: (4) genes with overlapping gene expression levels at some time points cannot be selected, and (5) using multiple prediction results at each time point is difficult.

To solve the problem of eENCD, we developed a CPMTP. CPMTP uses the difference in gene expression levels between one and the subsequent time points. Thus, CPMTP can select genes with different time-course gene expression levels overlapping at some time points between the two groups. Another point of CPMTP was that Bayesian theory was used to integrate multiple prediction results at each time point. This integration makes a prediction using time-course gene expression profiles easy. According to previous studies,

in HCV and MS datasets, CPMTP could predict therapeutic response with higher predictive performance than the conventional method and select the genes related to HCV and MS. Moreover, it was found that an early prediction point was essential for accurate prediction.

The eENCD and CPMTP were compared based on the predictive performance and function of genes selected by each method. This evaluation used time-course gene expression profiles of patients with MS. The results showed that the predictive performance of CPMTP was higher than that of eENCD. In addition, CPMTP could select more genes related to MS than eENCD. Thus, CPMTP was considered superior to eENCD.

The eENCD and CPMTP exhibit potential as useful biomarker discovery methods. Future work should validate these methods using data from more patients and other diseases. Validation with cancer data for which several biomarkers have been identified is helpful for accurately evaluating the performance of genes selected by eENCD and CPMTP as biomarkers. Another direction for future work is establishing a pre-processing method for microarray data, such as normalization. Variations in gene expression levels across microarrays can affect the performance of biomarkers, but variations in microarrays are one of the main issues that are difficult to solve. Moreover, to allow the genes to be selected by eENCD or CPMTP as clinical biomarkers, a biomarker discovery method using computers and experiments both *in vivo* and *in vitro* is essential.

In this study, eENCD and CPMTP were developed and evaluated. The proposed methods were found to predict therapeutic responses accurately and select genes related to the target of diseases from a massive gene in time-course expression profiles. Although

many validations are required for these genes to become clinical biomarkers, eENCD and CPMTP help develop new therapeutic strategies quickly and help reduce costs and time for development by providing candidate biomarker genes. In addition, it was suggested that early time points were essential for the accurate prediction of therapeutic responses through the evaluation of CPMTP. This suggestion helps develop clinical biomarkers from the perspective of determining time points used in *in vivo/in vitro* experiments. Therefore, eENCD and CPMTP can contribute to the development of new therapeutic strategies, such as personalized medicine using biomarkers.

References

[1] K. Stankov, "Bioinformatic tools for cancer geneticists," *Arch. Oncol.*, vol. 13, no. 2, pp. 69–75, 2005.

[2] A. M. Glazier, J. H. Nadeau, and T. J. Aitman, "Finding genes that underlie complex traits," *Science*, vol. 298, no. 5602, pp. 2345–2349, 2002.

[3] P. A. Futreal, A. Kasprzyk, E. Birney, J. C. Mullikin, R. Wooster, and M. R. Stratton, "Cancer and genomics," *Nature*, vol. 409, no. 6822, pp. 850–852, 2001.

[4] K. E. Peters, C. C. Walters, and J. M. Moldowan, "The Biomarker Guide," *Cambridge University Press*, 2004.

[5] N. Segata, J. Izard, L. Waldron, D. Gevers, L. Miropolsky, W. S. Garrett, and C. Huttenhower, "Metagenomic biomarker discovery and explanation," *Genome Biol.*, vol. 12, no. 6, p. R60, 2011.

[6] C. L. Sawyers, "The cancer biomarker problem," *Nature*, vol. 452, no. 7187, pp. 548–552, 2008.

[7] D. Sarker and P. Workman, "Pharmacodynamic biomarkers for molecular cancer therapeutics," *Adv. Cancer Res.*, vol. 96, pp. 213–268, 2007.

[8] C. E. Redon, A. J. Nakamura, Y. Zhang, J. J. Ji, W. M. Bonner, Robert J Kinders, R. E. Parchment, J. H. Doroshow, and Y. Pommier, "Histone gammaH2AX and poly(ADP-ribose) as clinical pharmacodynamic biomarkers," Clin. Cancer Res., vol. 16, no. 18, pp. 4532–4542, 2010.

[9] K. V. Ballman, "Biomarker: Predictive or Prognostic?," *J. Clin. Oncol.*, vol. 33, no. 33, pp. 3968–3971, 2015.

[10] M. B. Flanagan, D. J. Dabbs, A. M. Brufsky, S. Beriwal, and R. Bhargava,

"Histopathologic variables predict Oncotype DX recurrence score," *Mod. Pathol.*, vol. 21, no. 10, pp. 1255–1261, 2008.

[11] S. P. Patel and R. Kurzrock, "PD-L1 Expression as a Predictive Biomarker in Cancer Immunotherapy," *Mol. Cancer Ther.*, vol. 14, no. 4, pp. 847–856, 2015.

[12] E. A. Mittendorf, A. V. Philips, F. Meric-Bernstam, N. Qiao, Y. Wu, S. Harrington, X. Su, Y. Wang, A. M. Gonzalez-Angulo, A. Akcakanat, A. Chawla, M. Curran, P. Hwu, P. Sharma, J. K. Litton, J. J. Molldrem, and Gheath Alatrash, "PD-L1 expression in triple-negative breast cancer," *Cancer Immunol. Res.*, vol. 2, no. 4, pp. 361–370, 2014.

[13] M. J. Heller, "DNA microarray technology: devices, systems, and applications," *Annu. Rev. Biomed. Eng.*, vol. 4, pp. 129–153, 2002.

[14] D. Shalon, S. J. Smith, and P. O. Brown, "A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization," *Genome Res.*, vol. 6, no. 7, pp. 639–645, 1996.

[15] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, "Quantitative monitoring of gene expression patterns with a complementary DNA microarray," *Science*, vol. 270, no. 5235, pp. 467–470, 1995.

[16] N. Dorrell, J. A. Mangan, K. G. Laing, J. Hinds, D. Linton, H. Al-Ghusein, B. G. Barrell, J. Parkhill, N. G. Stoker, A. V. Karlyshev, P. D. Butcher, and B. W. Wren, "Whole genome comparison of Campylobacter jejuni human isolates using a low-cost microarray reveals extensive genetic diversity," *Genome Res.*, vol. 11, no. 10, pp. 1706–1715, 2001.

[17] Y. Wang, C. Barbacioru, F. Hyland, W. Xiao, K. L. Hunkapiller, J. Blake, F. Chan,

C. Gonzalez, L. Zhang, and R. R. Samaha, “Large scale real-time PCR validation on gene expression measurements from two commercial long-oligonucleotide microarrays,” *BMC Genomics*, vol. 7, no. 1, p. 59, 2006.

[18] R. Nadon and J. Shoemaker, “Statistical issues with microarrays: processing and analysis,” *Trends Genet.*, vol. 18, no. 5, pp. 265–71, 2002.

[19] B. Mei and Z. Wang, “An efficient method to handle the ‘large p, small n’ problem for genomewide association studies using Haseman-Elston regression,” *J. Genet.*, vol. 95, no. 4, pp. 847–852, 2016.

[20] D. B. Allison, X. Cui, G. P. Page, and M. Sabripour, “Microarray data analysis: from disarray to consolidation and consensus,” *Nat. Rev. Genet.*, vol. 7, no. 1, pp. 55–65, 2006.

[21] C. Isella, T. Renzulli, D. Corà, and E. Medico, “Mulcom: a multiple comparison statistical test for microarray data in Bioconductor,” *BMC Bioinformatics*, vol. 12, p. 382, 2011.

[22] Y. Ge, S. Dudoit, and T. P. Speed, “Resampling-based multiple testing for microarray data analysis,” *Test*, vol. 12, no. 1, pp. 1–77, 2003.

[23] T. Nichols and S. Hayasaka, “Controlling the familywise error rate in functional neuroimaging: a comparative review,” *Stat. Methods Med. Res.*, vol. 12, no. 5, pp. 419–446, 2003.

[24] Y. Benjamini and D. Yekutieli, “The control of the false discovery rate in multiple testing under dependency,” *Ann. Stat.*, vol. 29, no. 4, pp. 1165-1188, 2001.

[25] T. V. Perneger, “What’s wrong with Bonferroni adjustments,” *BMJ*, vol. 316, no. 7139, pp. 1236–1238, 1998.

[26] M. Aickin and H. Gensler, "Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods," *Am. J. Public Health*, vol. 86, no. 5, pp. 726–728, 1996.

[27] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *J. R. Stat. Soc. Ser. B*, vol. 57, no. 1, pp. 289–300, 1995.

[28] R. L. Somorjai, B. Dolenko, and R. Baumgartner, "Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions," *Bioinformatics*, vol. 19, no. 12, pp. 1484–1491, 2003.

[29] P. Indyk and R. Motwani, "Approximate nearest neighbors," *in Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pp. 604–613, 1998

[30] S. Debey, T. Zander, B. Brors, A. Popov, R. Eils, and J. L. Schultze, "A highly standardized, robust, and cost-effective method for genome-wide transcriptome analysis of peripheral blood applicable to large-scale clinical trials," *Genomics*, vol. 87, no. 5, pp. 653–664, 2006.

[31] C. M. Carvalho, J. Chang, J. E. Lucas, J. R. Nevins, Q. Wang, and M. West, "High-Dimensional Sparse Factor Modeling: Applications in Gene Expression Genomics," *J. Am. Stat. Assoc.*, vol. 103, no. 484, pp. 1438–1456, 2008.

[32] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A Sparse-Group Lasso," *J. Comput. Graph. Stat.*, vol. 22, no. 2, pp. 231–245, 2013.

[33] J. Ye, M. Farnum, E. Yang, R. Verbeeck, V. Lobanov, N. Raghavan, G. Novak, A. DiBernardo, and V. A. Narayan, "Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data," *BMC Neurol.*, vol.

12, no. 1, p. 46, 2012.

[34] A. Fukushima, M. Sugimoto, S. Hiwa, and T. Hiroyasu, “Elastic net-based prediction of IFN-β treatment response of patients with multiple sclerosis using time series microarray gene expression profiles,” *Sci. Rep.*, vol. 9, no. 1, p. 1822, 2019.

[35] R. Tishbirani, “Regression shrinkage and selection via the Lasso,” *J. R. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[36] W. Li and Y. Yang, “How Many Genes are Needed for a Discriminant Microarray Data Analysis,” *Methods of Microarray Data Analysis*, pp. 137–149, 2002.

[37] M. Xiong, X. Fang, and J. Zhao, “Biomarker identification by feature wrappers,” *Genome Res.*, vol. 11, no. 11, pp. 1878–1887, 2001.

[38] A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn, and A. A. Alizadeh, “Robust enumeration of cell subsets from tissue expression profiles,” *Nat. Methods*, vol. 12, no. 5, pp. 453–457, 2015.

[39] D. E. Farrar and R. R. Glauber, “Multicollinearity in Regression Analysis: The Problem Revisited,” *Rev. Econ. Stat.*, vol. 49, no. 1, p. 92, 1967.

[40] A. Senawi, H.-L. Wei, and S. A. Billings, “A new maximum relevance-minimum multicollinearity (MRmMC) method for feature selection and ranking,” *Pattern Recognit.*, vol. 67, pp. 47–61, 2017.

[41] C. Ding and H. Peng, “Minimum redundancy feature selection from microarray gene expression data,” *J. Bioinform. Comput. Biol.*, vol. 3, no. 2, pp. 185–205, 2005.

[42] A. Garg and K. Tai, “Comparison of statistical and machine learning methods in

modelling of data with multicollinearity," *Int. J. Model. Identif. Control*, vol. 18, no. 4, p. 295, 2013.

[43] D. Lin, D. P. Foster, and L. H. Ungar, "VIF Regression: A Fast Regression Algorithm for Large Data," *J. Am. Stat. Assoc.*, vol. 106, no. 493, pp. 232–247, 2011.

[44] K. Horimoto and H. Toh, "Statistical estimation of cluster boundaries in gene expression profile data," *Bioinformatics*, vol. 17, no. 12, pp. 1143–1151, 2001.

[45] Z. Y. Algamal and M. H. Lee, "Adjusted Adaptive LASSO in High-dimensional Poisson Regression Model," *Mod. Appl. Sci.*, vol. 9, no. 4, pp. 170-177, 2015.

[46] I.-G. Chong and C.-H. Jun, "Performance of some variable selection methods when multicollinearity is present," *Chemom. Intell. Lab. Syst.*, vol. 78, no. 1–2, pp. 103–112, 2005.

[47] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc. Ser. B (Statistical Methodol.)*, vol. 67, no. 5, pp. 768–768, 2005.

[48] L. L. Elo and B. Schwikowski, "Analysis of time-resolved gene expression measurements across individuals," *PLoS One*, vol. 8, no. 12, p. e82340, 2013.

[49] Z. Bar-Joseph, A. Gitter, and I. Simon, "Studying and modelling dynamic biological processes using time-series gene expression data," *Nat. Rev. Genet.*, vol. 13, no. 8, pp. 552–564, 2012.

[50] M. Kayano, H. Matsui, R. Yamaguchi, S. Imoto, and S. Miyano, "Gene set differential analysis of time course expression profiles via sparse estimation in functional logistic model with application to time-dependent biomarker detection," *Biostatistics*, vol. 17, no. 2, pp. 235–248, 2016.

[51] J. H. Phan and M. D. Wang, "Estimating Classification Error to Identify Biomarkers in Time Series Expression Data," *in Proceedings of IEEE 7th International Symposium on BioInformatics and BioEngineering*, pp. 172–179, 2007.

[52] A. Papana and H. Ishwaran, "Gene hunting with forests for multigroup time course data," *Stat. Probab. Lett.*, vol. 79, no. 9, pp. 1146–1154, 2009.

[53] T. Barrett, D. B. Troup, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, R. N. Muertter, M. Holko, O. Ayanbule, A. Yefanov, and A. Soboleva, "NCBI GEO: archive for functional genomics data sets--10 years on," *Nucleic Acids Res.*, vol. 39, pp. D1005-1010, 2011.

[54] Z. R. Siow, R. H. De Boer, G. J. Lindeman, and G. B. Mann, "Spotlight on the utility of the Oncotype DX® breast cancer assay," *Int. J. Womens. Health*, vol. 10, pp. 89–100, 2018.

[55] J. J. Carlson and J. A. Roth, "The impact of the Oncotype Dx breast cancer assay in clinical practice: a systematic review and meta-analysis," *Breast Cancer Res. Treat.*, vol. 141, no. 1, pp. 13–22, 2013.

[56] R. Liu, X. Wang, K. Aihara, and L. Chen, "Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers," *Med. Res. Rev.*, vol. 34, no. 3, pp. 455–478, 2014.

[57] L. Chen, R. Liu, Z.-P. Liu, M. Li, and K. Aihara, "Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers," *Sci. Rep.*, vol. 2, no. 1, p. 342, 2012.

[58] T. Huang, K. Tu, Y. Shyr, C.-C. Wei, L. Xie, and Y.-X. Li, "The prediction of interferon treatment effects based on time series microarray gene expression profiles," *J. Transl. Med.*, vol. 6, p. 44, 2008.

[59] H. Wang, H. Sun, T. Chang, H. Lo, W. Cheng, G. C. Tseng, C. Lin, S. Chang, N. Pal, and I. Chung, "Discovering monotonic stemness marker genes from time-series stem cell microarray data," *BMC Genomics*, vol. 16, no. S2, pp. 1-18, 2015.

[60] X. Leng and H.-G. Müller, "Classification using functional data analysis for temporal gene expression data," *Bioinformatics*, vol. 22, no. 1, pp. 68–76, 2006.

[61] B. Di Camillo, G. Toffolo, S. K. Nair, L. J. Greenlund, and C. Cobelli, "Significance analysis of microarray transcript levels in time series experiments," *BMC Bioinformatics*, vol. 8, no. S10, pp. 1-13, 2007.

[62] D. Rizopoulos, G. Molenberghs, and E. M. E. H. Lesaffre, "Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking," *Biom. J.*, vol. 59, no. 6, pp. 1261–1276, 2017.

[63] K. Li and S. Luo, "Dynamic predictions in Bayesian functional joint models for longitudinal and time-to-event data: An application to Alzheimer's disease," *Stat. Methods Med. Res.*, vol. 28, no. 2, pp. 327-342, 2019.

[64] B. Hemmer, J. J. Archelos, and H.-P. Hartung, "New concepts in the immunopathogenesis of multiple sclerosis," *Nat. Rev. Neurosci.*, vol. 3, no. 4, pp. 291–301, 2002.

[65] J. Howard, S. Trevick, and D. S. Younger, "Epidemiology of Multiple Sclerosis," *Neurol. Clin.*, vol. 34, no. 4, pp. 919–939, 2016.

[66] S. L. Hauser, E. Waubant, D. L. Arnold, T. Vollmer, J. Antel, R. J. Fox, A. Bar-Or,

M. Panzara, N. Sarkar, S. Agarwal, A. Langer-Gould, and C. H. Smith, "B-cell depletion with rituximab in relapsing-remitting multiple sclerosis," *N. Engl. J. Med.*, vol. 358, no. 7, pp. 676–688, 2008.

[67] M. K. Singh, T. F. Scott, W. A. LaFramboise, F. Z. Hu, J. C. Post, and G. D. Ehrlich, "Gene expression changes in peripheral blood mononuclear cells from multiple sclerosis patients undergoing beta-interferon therapy," *J. Neurol. Sci*., vol. 258, no. 1–2, pp. 52–59, 2007.

[68] S. E Baranzini, P. Mousavi, J. Rio, S. J. Caillier, A. Stillman, P. Villoslada, M. M. Wyatt, M. Comabella, L. D. Greller, R. Somogyi, X. Montalban, and J. R. Oksenberg, "Transcription-based prediction of response to IFNbeta using supervised computational methods," *PLoS Biol.*, vol. 3, no. 1, p. e2, 2005.

[69] R. A. Rudick, M. R. S. Rani, Y. Xu, J. Lee, J. Na, J. Shrock, A. Josyula, E. Fisher, and R. M. Ransohoff, "Excessive biologic response to IFNβ is associated with poor treatment response in patients with multiple sclerosis," *PLoS One*, vol. 6, no. 5, p. e19262, 2011.

[70] J. Río, C. Nos, M. Tintoré, N. Téllez, I. Galán, R. Pelayo, M. Comabella, and X. Montalban, "Defining the response to interferon-beta in relapsing-remitting multiple sclerosis patients," *Ann. Neurol.*, vol. 59, no. 2, pp. 344–352, 2006.

[71] A. Hundeshagen, M. Hecker, B. K. Paap, C. Angerstein, O. Kandulski, C. Fatum, C. Hartmann, D. Koczan, H. Thiesen, and U. K. Zettl, "Elevated type I interferon-like activity in a subset of multiple sclerosis patients: molecular basis and clinical relevance," *J. Neuroinflammation*, vol. 9, no. 1, p. 140, 2012.

[72] S. Malhotra, M. F. Bustamante, F. Pérez-Miralles, J. Rio, M. C. R. Villa, E. Vegas,

L. Nonell, F. Deisenhammer, N. Fissolo, R. N. Nurtdinov, X. Montalban, and M. Comabella, “Search for specific biomarkers of IFNβ bioactivity in patients with multiple sclerosis,” *PLoS One*, vol. 6, no. 8, p. e23634, 2011.

[73] F. Gilli, F. Marnetto, M. Caldano, A. Sala, S. Malucchi, M. Capobianco, and A. Bertolotto, “Biological markers of interferon-beta therapy: comparison among interferon-stimulated genes MxA, TRAIL and XAF-1,” *Mult. Scler.*, vol. 12, no. 1, pp. 47–57, 2006.

[74] M. Hecker, B. K. Paap, R. H. Goertsches, O. Kandulski, C. Fatum, D. Koczan, H. Hartung, H. Thiesen, and U. K. Zettl, “Reassessment of blood gene expression markers for the prognosis of relapsing-remitting multiple sclerosis,” *PLoS One*, vol. 6, no. 12, p. e29648, 2011.

[75] [S. Martire, N. D. Navone, F. Montarolo, S. Perga, and A. Bertolotto, “A gene expression study denies the ability of 25 candidate biomarkers to predict the interferon-beta treatment response in multiple sclerosis patients,” *J. Neuroimmunol.*, vol. 292, pp. 34–9, 2016.

[76] K. Morino, Y. Hirata, R. Tomioka, H. Kashima, K. Yamanishi, N. Hayashi, S. Egawa, and K. Aihara, “Predicting disease progression from short biomarker series using expert advice algorithm,” *Sci. Rep.*, vol. 5, p. 8953, 2015.

[77] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization Paths for Generalized Linear Models via Coordinate Descent,” *J. Stat. Softw.*, vol. 33, no. 1, pp. 1–22, 2010.

[78] L. Meier, S. Van De Geer, and P. Bühlmann, “The group lasso for logistic regression,” *J. R. Stat. Soc. Ser. B (Statistical Methodol.)*, vol. 70, no. 1, pp. 53–

71, 2008.

[79] J. J. Hughey and A. J. Butte, “Robust meta-analysis of gene expression using the elastic net,” *Nucleic Acids Res.*, vol. 43, no. 12, p. e79, 2015.

[80] J. Fan and J. Lv, “Sure independence screening for ultrahigh dimensional feature space,” *J. R. Stat. Soc. Ser. B (Statistical Methodol.*), vol. 70, no. 5, pp. 849–911, 2008.

[81] M.-Y. Wu, X.-F. Zhang, D.-Q. Dai, L. Ou-Yang, Y. Zhu, and H. Yan, “Regularized logistic regression with network-based pairwise interaction for biomarker identification in breast cancer,” *BMC Bioinformatics*, vol. 17, p. 108, 2016.

[82] T. Shimamura, S. Imoto, R. Yamaguchi, A. Fujita, M. Nagasaki, and S. Miyano, “Recursive regularization for inferring gene networks from time-course gene expression profiles,” *BMC Syst. Biol.*, vol. 3, p. 41, 2009.

[83] N. Meinshausen and P. Bühlmann, “Stability selection,” *J. R. Stat. Soc. Ser. B (Statistical Methodol.)*, vol. 72, no. 4, pp. 417–473, 2010.

[84] M. Hecker, R. H. Goertsches, C. Fatum, D. Koczan, H. Thiesen, R. Guthke, and U. K. Zettl, “Network analysis of transcriptional regulation in response to intramuscular interferon-β-1a multiple sclerosis treatment,” *Pharmacogenomics J.*, vol. 12, no. 2, pp. 134–146, 2012.

[85] R. H. Goertsches, M. Hecker, D. Koczan, P. Serrano-Fernandez, S. Moeller, H. Thiesen, and U. K. Zettl, “Long-term genome-wide blood RNA expression profiles yield novel molecular response candidates for IFN-beta-1b treatment in relapsing remitting MS,” *Pharmacogenomics*, vol. 11, no. 2, pp. 147–161, 2010.

[86] R. Kohavi, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation

and Model Selection," *in Proceedings of International Joint Conference of Artificial Intelligence*, pp. 1137–1143, 1995.

[87] M. Tsagris, V. Lagani, and I. Tsamardinos, "Feature selection for high-dimensional temporal data," *BMC Bioinformatics*, vol. 19, no. 1, p. 17, 2018.

[88] S. P. Singh, H. H. Zhang, H. Tsang, P. J. Gardina, T. G. Myers, V. Nagarajan, C. H. Lee, and J. M. Farber, "PLZF regulates CCR6 and is critical for the acquisition and maintenance of the Th17 phenotype in human cells," *J. Immunol.*, vol. 194, no. 9, pp. 4350–4361, 2015.

[89] A. Jones and D. Hawiger, "Peripherally Induced Regulatory T Cells: Recruited Protectors of the Central Nervous System against Autoimmune Neuroinflammation," *Front. Immunol.*, vol. 8, p. 532, 2017.

[90] A. Jones, A. Opejin, J. G. Henderson, C. Gross, R. Jain, J. A. Epstein, R. A. Flavell, and D. Hawiger, "Peripherally Induced Tolerance Depends on Peripheral Regulatory T Cells That Require Hopx To Inhibit Intrinsic IL-2 Expression," *J. Immunol.*, vol. 195, no. 4, pp. 1489–1497, 2015.

[91] K. Conant, S. Daniele, P. L. Bozzelli, T. Abdi, A. Edwards, A. Szklarczyk, I. Olchefske, D. Ottenheimer, and K. Maguire-Zeiss, "Matrix metalloproteinase activity stimulates N-cadherin shedding and the soluble N-cadherin ectodomain promotes classical microglial activation," *J. Neuroinflammation*, vol. 14, no. 1, p. 56, 2017.

[92] A. F. Leuchter, I. A. Cook, S. P. Hamilton, K. L. Narr, A. Toga, A. M. Hunter, K. Faull, J. Whitelegge, A. M. Andrews, J. Loo, B. Way, S. F. Nelson, S. Horvath, and B. D. Lebowitz, "Biomarkers to predict antidepressant response," *Curr.*

Psychiatry Rep., vol. 12, no. 6, pp. 553–562, 2010.

[93] M. Oswald, M. E. Curran, S. L. Lamberth, R. M. Townsend, J. D. Hamilton, D. N. Chernoff, J. Carulli, M. J. Townsend, M. E. Weinblatt, M. Kern, C. M. Pond, A. Lee, and P. K. Gregersen, "Modular analysis of peripheral blood gene expression in rheumatoid arthritis captures reproducible gene expression changes in tumor necrosis factor responders," *Arthritis Rheumatol. (Hoboken, N.J.)*, vol. 67, no. 2, pp. 344–351, 2015.

[94] H. Hofer, T. Watkins-Riedel, O. Janata, E. Penner, H. Holzmann, P. Steindl-Munda, A. Gangl, and P. Ferenci, "Spontaneous viral clearance in patients with acute hepatitis C can be predicted by repeated measurements of serum viral load," *Hepatology*, vol. 37, no. 1, pp. 60–64, 2003.

[95] C. C. Crone, G. M. Gabriel, and T. N. Wise, "Managing the neuropsychiatric side effects of interferon-based therapy for hepatitis C," *Cleve. Clin. J. Med.*, vol. 71, no. SUPPL. 3, pp. 27–32, 2004.

[96] X. Huang, J. Zeng, L. Zhou, C. Hu, P. Yin, and X. Lin, "A New Strategy for Analyzing Time-Series Data Using Dynamic Networks: Identifying Prospective Biomarkers of Hepatocellular Carcinoma," *Sci. Rep.*, vol. 6, p. 32448, 2016.

[97] A. Machireddy, G. Thibault, A. Tudorica, A. Afzal, M. Mishal, K. Kemmer, A. Naik, M. Troxell, E. Goranson, K. Oh, N. Roy, N. Jafarian, M. Holtorf, W. Huang, and X. Song, "Early Prediction of Breast Cancer Therapy Response using Multiresolution Fractal Analysis of DCE-MRI Parametric Maps," *Tomogr. (Ann Arbor, Mich.)*, vol. 5, no. 1, pp. 90–98, 2019.

[98] A. Gordon, G. Glazko, X. Qiu, and A. Yakovlev, "Control of the mean number of

false discoveries, Bonferroni and stability of multiple testing," *Ann. Appl. Stat.*, vol. 1, no. 1, pp. 179–190, 2007.

[99] Y. J. Bae et al., "Time-course microarray analysis for identifying candidate genes involved in obesity-associated pathological changes in the mouse colon," *Genes Nutr.*, vol. 11, no. 1, p. 30, 2016.

[100] Y. J. Bae, S. Kim, S. Y. Hong, T. Park, S. Lee, M. Choi, and M. Sung, "Analysis of Genes Related to Angiotensin II-Induced Arterial Injury Using a Time Series Microarray," *Cell. Physiol. Biochem.*, vol. 48, no. 3, pp. 983–992, 2018.

[101] D. Rizopoulos, L. A. Hatfield, B. P. Carlin, and J. J. M. Takkenberg, "Combining Dynamic Predictions From Joint Models for Longitudinal and Time-to-Event Data Using Bayesian Model Averaging," *J. Am. Stat. Assoc.*, vol. 109, no. 508, pp. 1385–1397, 2014.

[102] K. Y. Yeung, R. E. Bumgarner, and A. E. Raftery, "Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data," *Bioinformatics*, vol. 21, no. 10, pp. 2394–2402, 2005.

[103] A. Annest, R. E. Bumgarner, A. E. Raftery, and K. Y. Yeung, "Iterative Bayesian Model Averaging: a method for the application of survival analysis to high-dimensional microarray data," *BMC Bioinformatics*, vol. 10, p. 72, 2009.

[104] D. P. Berrar, C. S. Downes, and W. Dubitzky, "Multiclass cancer classification using gene expression profiling and probabilistic neural networks," *Pac. Symp. Biocomput.*, pp. 5–16, 2003.

[105] Z. M. Hira and D. F. Gillies, "A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data," *Adv. Bioinformatics*, vol. 2015,

no. 1, p. 198363, 2015.

[106] A. Conesa, M. J. Nueda, A. Ferrer, and M. Talón, "maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments," *Bioinformatics*, vol. 22, no. 9, pp. 1096–1102, 2006.

[107] S. Ma, X. Song, and J. Huang, "Supervised group Lasso with applications to microarray data analysis," *BMC Bioinformatics*, vol. 8, p. 60, 2007.

[108] M. W. Taylor, T. Tsukahara, L. Brodsky, J. Schaley, C. Sanda, M. J. Stephens, J. N. McClintick, H. J. Edenberg, L. Li, J. E. Tavis, C. Howell, and S. H. Belle, "Changes in gene expression during pegylated interferon and ribavirin therapy of chronic hepatitis C virus distinguish responders from nonresponders to antiviral therapy," *J. Virol.*, vol. 81, no. 7, pp. 3391–3401, 2007.

[109] N. Q. Tran, M. koçak, and M. Mendeş, "Comparison of Commonly Used Methods for Testing Interaction Effect in Time-Course Microarray Experiments," *Turkiye Klin. J. Biostat.*, vol. 9, no. 1, pp. 35–44, 2017.

[110] L. Yin, Y. Wang, Y. Lin, G. Yu, and Q. Xia, "Explorative analysis of the gene expression profile during liver regeneration of mouse: a microarray-based study," *Artif. cells, nanomedicine, Biotechnol.*, vol. 47, no. 1, pp. 1113–1121, 2019.

[111] L.-Y. Chuang, C.-H. Yang, J.-C. Li, and C.-H. Yang, "A Hybrid BPSO-CGA Approach for Gene Selection and Classification of Microarray Data," *J. Comput. Biol.*, vol. 19, no. 1, pp. 68–82, 2012.

[112] H. Motieghader, A. Najafi, B. Sadeghi, and A. Masoudi-Nejad, "A hybrid gene selection algorithm for microarray cancer classification using genetic algorithm and learning automata," *Informatics Med. Unlocked*, vol. 9, pp. 246–254, 2017.

[113] E. E. Powell, C. J. Edwards-Smith, J. L. Hay, A. D. Clouston, D. H. Crawford, C. Shorthouse, D. M. Purdie, and J. R. Jonsson, "Host genetic factors influence disease progression in chronic hepatitis C," *Hepatology*, vol. 31, no. 4, pp. 828–833, 2000.

[114] M. W. Fried, M. L. Shiffman, K. R. Reddy, C. Smith, G. Marinos, F. L. Gonçales Jr, D. Häussinger, M. Diago, G. Carosi, D. Dhumeaux, A. Craxi, A. Lin, J. Hoffman, and J. Yu, "Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection," *N. Engl. J. Med.*, vol. 347, no. 13, pp. 975–982, 2002.

[115] T. Kuwashiro, S. Iwane, X. Jinghe, S. Matsuhash, Y. Eguchi, K. Anzai, K. Fujimoto, T. Mizuta, N. Sakamoto, M. Ikeda, N. Kato, and I. Ozaki, "Regulation of interferon signaling and HCV-RNA replication by extracellular matrix," *Int. J. Mol. Med.*, vol. 42, no. 2, pp. 957–965, 2018.

[116] S. L. Friedman, "Mechanisms of hepatic fibrogenesis," *Gastroenterology*, vol. 134, no. 6, pp. 1655–69, May 2008.

[117] Y. Kushima, T. Wakita, and M. Hijikata, "A disulfide-bonded dimer of the core protein of hepatitis C virus is important for virus-like particle production," *J. Virol.*, vol. 84, no. 18, pp. 9118–9127, 2010.

[118] L. P. Shriver and M. Manchester, "Inhibition of fatty acid metabolism ameliorates disease activity in an animal model of multiple sclerosis," *Sci. Rep.*, vol. 1, p. 79, 2011.

[119] M. Bock, M. Karber, and H. Kuhn, "Ketogenic diets attenuate cyclooxygenase and lipoxygenase gene expression in multiple sclerosis," *EBioMedicine*, vol. 36, pp. 293–303, 2018.

[120] S. Marusic, P. Thakker, J. W. Pelker, N. L. Stedman, K. L. Lee, J. C. McKew, L. Han, X. Xu, S. F. Wolf, A. J. Borey, J. Cui, M. W. H Shen, F. Donahue, M. Hassan-Zahraee, Michael W. Leach, T. Shimizu, and J. D. Clark, "Blockade of cytosolic phospholipase A2 alpha prevents experimental autoimmune encephalomyelitis and diminishes development of Th1 and Th17 responses," *J. Neuroimmunol.*, vol. 204, no. 1–2, pp. 29–37, 2008.

[121] T. Udhayabanu, A. Manole, M. Rajeshwari, P. Varalakshmi, H. Houlden, and B. Ashokkumar, "Riboflavin Responsive Mitochondrial Dysfunction in Neurodegenerative Diseases," *J. Clin. Med.*, vol. 6, no. 5, p. 52, 2017.

[122] T. Sakurai, S. Miyazawa, S. Furuta, and T. Hashimoto, "Riboflavin deficiency and beta-oxidation systems in rat liver," *Lipids*, vol. 17, no. 9, pp. 598–604, 1982.

[123] M. Naghashpour, S. Jafarirad, R. Amani, A. Sarkaki, and A. Saedisomeolia, "Update on riboflavin and multiple sclerosis: a systematic review," *Iran. J. Basic Med. Sci.*, vol. 20, no. 9, pp. 958–966, 2017.

[124] R. Dutta, J. McDonough, and X. Yin *et al.*, "Mitochondrial dysfunction as a cause of axonal degeneration in multiple sclerosis patients," *Annals of Neurology*, vol. 59, no. 3, pp. 478–489, 2006.

[125] H. R. Scholte, H. F. Busch, H. D. Bakker, J. M. Bogaard, I. E. Luyt-Houwen, and L. P. Kuyt, "Riboflavin-responsive complex I deficiency," *Biochim. Biophys. Acta*, vol. 1271, no. 1, pp. 75–83, 1995.

[126] M. Gerards, B. J. C. Bosch, K. Danhauser, V. Serre, M. Weeghel, R. J. A. Wanders, G. A. F. Nicolaes, W. Sluiter, K. Schoonderwoerd, H. R. Scholte, H. Prokisch, A. Rötig, I. F. M. Coo, and H. J. M. Smeets, "Riboflavin-responsive oxidative

phosphorylation complex I deficiency caused by defective ACAD9: new function for an old gene," *Brain*, vol. 134, no. Pt 1, pp. 210–219, 2011.

[127] C. Garone, M. A. Donati, M. Sacchini, B. Garcia-Diaz, C. Bruno, S. Calvo, V. K. Mootha, and S. Dimauro, "Mitochondrial encephalomyopathy due to a novel mutation in ACAD9," *JAMA Neurol.*, vol. 70, no. 9, pp. 1177–1179, 2013.

[128] S. C. Cunnane, S. Y. Ho, P. Dore-Duffy, K. R. Ells, and D. F. Horrobin, "Essential fatty acid and lipid profiles in plasma and erythrocytes in patients with multiple sclerosis," *Am. J. Clin. Nutr.*, vol. 50, no. 4, pp. 801–806, 1989.

[129] H.-H. Kim, I. H. Jeong, J.-S. Hyun, B. S. Kong, H. J. Kim, and S. J. Park, "Metabolomic profiling of CSF in multiple sclerosis and neuromyelitis optica spectrum disorder by nuclear magnetic resonance," *PLoS One*, vol. 12, no. 7, p. e0181758, 2017.

[130] G. S. M. Ramsaransing, M. R. Fokkema, A. Teelken, A. V Arutjunyan, M. Koch, and J. De Keyser, "Plasma homocysteine levels in multiple sclerosis," *J. Neurol. Neurosurg. Psychiatry*, vol. 77, no. 2, pp. 189–192, 2006.

[131] D. A Monti, G. Zabrecky, T. P. Leist, N. Wintering, A. J. Bazzan, T. Zhan, and A. B. Newberg, "N-acetyl Cysteine Administration Is Associated With Increased Cerebral Glucose Metabolism in Patients With Multiple Sclerosis: An Exploratory Study," *Front. Neurol.*, vol. 11, p. 88, 2020.

[132] L. P. Shriver and M. Manchester, "Inhibition of fatty acid metabolism ameliorates disease activity in an animal model of multiple sclerosis," *Sci. Rep.*, vol. 1, p. 79, 2011.

Copyrights

The second and third sections are the reuse of "Elastic net-based prediction of IFN-β treatment response of patients with multiple sclerosis using time series microarray gene expression profiles (publisher: *Scientific Reports*, vol. 9, p. 1822)" and "Bayesian approach for predicting responses to therapy from high-dimensional time-course gene expression profiles (publisher: *BMC bioinformatics*, vol. 22, no. 1, p. 132)", respectively. These papers can be accessed from the following URLs.

The second section: https://www.nature.com/articles/s41598-018-38441-2

The third section:https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-021-04052-4

Publications

Journal articles:

1. 廣安知之, 福島亜梨花, 山本詩子, 横内久猛, “2 本の時系列データの類似部分自動抽出法の提案 —fNIRS 時系列データに対する検討”, 情報処理学会論文誌数理モデル化と応用, vol. 7, no. 2, pp. 64-73, 2014
2. A. Fukushima, T. Yano, S. Imahara, H. Aisu, Y. Shimokawa, Y. Shibata, “Prediction of energy consumption for new electric vehicle models by machine learning”, *IET Intelligent Transport Systems*, vol. 12, no.9, pp. 1174-1180. 2018.
3. A. Fukushima, M. Sugimoto, S. Hiwa, T. Hiroyasu, “Elastic net-based prediction of IFN-β treatment response of patients with multiple sclerosis using time series microarray gene expression profiles”, *Scientific Reports*, vol. 9, pp. 1822, 2019.
4. 高明淑, 福島亜梨花, 矢野亨, “決定木に基づく在宅リハビリ実施支援システムの提案－サンディング動作を対象とした検討－”, 電気学会論文誌Ｃ, vol. 139, no. 7, pp. 766-773, 2019.
5. 矢野亨, 福島亜梨花, 下川裕亮, 柴田康弘, 能登谷英樹, 田村聡一朗, “EV 充電ナビシステム実証実験による消費電力量予測の検証”, 電気学会論文誌Ｃ, vol. 140, no. 2, pp. 1174-1180, 2020.
6. A. Fukushima, M. Sugimoto, S. Hiwa, T. Hiroyasu, “Bayesian approach for predicting responses to therapy from high-dimensional time-course gene expression profiles”, *BMC bioinformatics*, vol. 22, no.1, p. 132, 2021.

International conference proceedings:

1. T. Hiroyasu, A. Fukushima, H. Yokouchi, “Differences in blood flow between auditory and visual stimuli in the Psychomotor Vigilance Task and GO/NOGO Task”, *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, Aug. 2012 (Poster presentation).
2. T. Hiroyasu, A. Fukushima, U. Yamamoto, “Extraction Algorithm of Similar Parts from Multiple Time-Series Data of Cerebral Blood Flow”, *Brain and Health Informatics*, p. 138-146, Maebashi, Oct. 2013 (Oral presentation).
3. A. Fukushima, T. Paul, R. Shingaki, T. Koiso, S. Umeno, K. Ueno, “A proposal for improvement of genotyping performance for ethnically homogeneous population using DNA microarray”, *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Milano, Aug. 2015 (Poster presentation).
4. A. Fukushima, T. Yano, S. Imahara, H. Aisu, Y. Shimokawa, Y. Shibata, “Prediction of energy consumption for new electric vehicle models by machine learning”, *ITS World Congress*, Copenhagen, Sep. 2018 (Oral presentation).
5. A. Fukushima, T. Yano, Y. Shimokawa, Y. Shibata, “Automatic construction of prediction models for energy consumption of various electric vehicles under various driving conditions”, *ITS World Congress*, Singapore, Oct. 2019 (Oral presentation).

Domestic conference proceedings (only first author):

1. 福島亜梨花, 廣安知之, 横内久猛, “PVT と GO/NOGO Task における視覚刺激と聴覚刺激に対する脳血流変化の違いの検討 福島亜梨花”, *日本生体医工学会大会*, 5 月 2012 (口頭発表).
2. 福島亜梨花, 廣安知之, 横内久猛, “脳血流時系列データの類似部分抽出、および神経活動の時間的遷移同定法の提案と検討”, *人工知能学会全国大会* , 5 月 2013 (口頭発表).
3. 福島亜梨花, 高明淑, 植野研, “L1 正則化を用いた時系列データ判別手法の提案”, *電子情報通信学会総合大会*, 3 月 2017 (口頭発表).
4. 福島亜梨花, 高明淑, 植野研, “動作フォームにおけるセンシング方法の検討”, *センシングフォーラム*, 8 月 2017 (口頭発表).
5. 福島亜梨花, 矢野亨, 今原修一郎, 愛須英之, “少数データに対する転移学習技術の提案と公開データによる評価”, *情報論的学習理論ワークショップ*, 東京, 11 月, 2017 (ポスター発表).
6. 福島亜梨花, 矢野亨, 山本幸弘, 愛須英之, 下川裕亮, 柴田康弘, “EV 充電ナビシステムのための消費電力量予測モデル自動構築技術”, *電気学会電子・情報・システム部門大会*, 北海道, 9 月 2018 (口頭発表).

7. 福島亜梨花, 矢野亨, 下川裕亮, 柴田康弘, 能登谷英樹, 田村聡一朗, “EV 充電ナビシステムの実走行データを用いた消費電力量予測モデル自動構築技術の評価”, *電気学会電子・情報・システム部門大会*, 沖縄, 9 月 2019 (口頭発表).

Awards:

1. ITS World Congress 2018 Best Scientific Paper Award (Asia-Pacific region), Sep. 2018.
2. 電気学会 電子・情報・システム部門大会 優秀論文発表賞, 9 月 2019.
3. 電気学会 電子・情報・システム部門大会 奨励賞, 9 月 2020.