



Universiteit
Leiden
The Netherlands

Identification of high-dimensional omics-derived predictors for tumor growth dynamics using machine learning and pharmacometric modeling

Zwep, L.B.; Duisters, K.L.W.; Jansen, M.; Guo, T.; Meulman, J.J.; Upadhyay, P.; Hasselt, J.G.C.

Citation

Zwep, L. B., Duisters, K. L. W., Jansen, M., Guo, T., Meulman, J. J., Upadhyay, P., & Hasselt, J. G. C. (2021). Identification of high-dimensional omics-derived predictors for tumor growth dynamics using machine learning and pharmacometric modeling. *Cpt: Pharmacometrics & Systems Pharmacology*, 10(4), 350-361. doi:10.1002/psp4.12603

Version: Publisher's Version

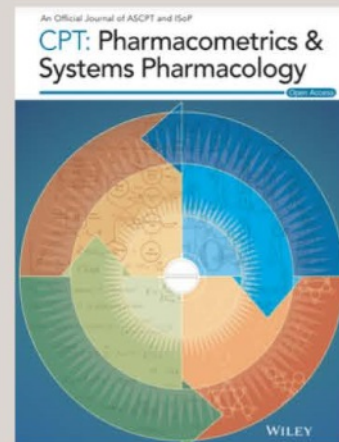
License: [Creative Commons CC BY-NC-ND 4.0 license](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3214540>

Note: To cite this publication please use the final published version (if applicable).

CPT: Pharmacometrics & Systems Pharmacology (PSP)

Accepted into Clarivate Analytics' SCIE (Science Citation Index Expanded) and received its **first Impact Factor of 4.054** in June 2021.



Why publish Open Access in *CPT: Pharmacometrics & Systems Pharmacology (PSP)*?

Easy submissions
with direct access to
editorial office
support

Open Access
official journal
of
ASCPT

Author support
with ASCPT tools
to amplify
research

Cross- disciplinary
publishing advances
in quantitative
methods

By choosing to publish Open Access with *CPT: Pharmacometrics & Systems Pharmacology (PSP)* your work will be made freely and globally available.

Publishing Open Access offers you many benefits — from reaching peers world-wide to securing more downloads and citations of your paper.



ARTICLE

Identification of high-dimensional omics-derived predictors for tumor growth dynamics using machine learning and pharmacometric modeling

Laura B. Zwep^{1,2} | Kevin L. W. Duisters² | Martijn Jansen¹ | Tingjie Guo^{1,3} |
Jacqueline J. Meulman² | Parth J. Upadhyay¹ | J. G. Coen van Hasselt¹

¹Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands

²Mathematical Institute, Leiden University, Leiden, The Netherlands

³Department of Intensive Care Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Correspondence

J. G. Coen van Hasselt and Laura B. Zwep, Division of Systems Biomedicine and Pharmacology, Leiden Academic Centre for Drug Research, Leiden University, Einsteinweg 55, 2333 CC, Leiden, The Netherlands.
Emails: coen.vanhasselt@lacdr.leidenuniv.nl, l.b.zwep@lacdr.leidenuniv.nl

Funding information

No funding was received for this work.

Abstract

Pharmacometric modeling can capture tumor growth inhibition (TGI) dynamics and variability. These approaches do not usually consider covariates in high-dimensional settings, whereas high-dimensional molecular profiling technologies (“omics”) are being increasingly considered for prediction of anticancer drug treatment response. Machine learning (ML) approaches have been applied to identify high-dimensional omics predictors for treatment outcome. Here, we aimed to combine TGI modeling and ML approaches for two distinct aims: omics-based prediction of tumor growth profiles and identification of pathways associated with treatment response and resistance. We propose a two-step approach combining ML using least absolute shrinkage and selection operator (LASSO) regression with pharmacometric modeling. We demonstrate our workflow using a previously published dataset consisting of 4706 tumor growth profiles of patient-derived xenograft (PDX) models treated with a variety of mono- and combination regimens. Pharmacometric TGI models were fit to the tumor growth profiles. The obtained empirical Bayes estimates-derived TGI parameter values were regressed using the LASSO on high-dimensional genomic copy number variation data, which contained over 20,000 variables. The predictive model was able to decrease median prediction error by 4% as compared with a model without any genomic information. A total of 74 pathways were identified as related to treatment response or resistance development by LASSO, of which part was verified by literature. In conclusion, we demonstrate how the combined use of ML and pharmacometric modeling can be used to gain pharmacological understanding in genomic factors driving variation in treatment response.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Pharmacometric tumor growth inhibition models represent a well-established approach—quantify tumor growth dynamics and the effects of therapeutics agents. High-dimensional molecular profiling technologies (“omics”) are relevant predictors for prediction of interindividual variation in tumor growth dynamics.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

WHAT QUESTION DID THIS STUDY ADDRESS?

How can high-dimensional omics-based predictors be efficiently identified for tumor growth inhibition model parameters?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

Two-step approach combining tumor growth inhibition modeling and least absolute shrinkage and selection operator regression allows identification of pharmacologically predictive and mechanistically relevant omics-derived molecular predictors for variation in tumor growth dynamics.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

This developed approach can be used to enable personalized omics-derived predictions of optimal treatments and dosing schedules, through its integration with tumor growth modeling. The integration with tumor growth modeling also allows identification of molecular predictors specifically associated with efficacy or treatment resistance.

INTRODUCTION

Pharmacometric modeling of tumor growth inhibition (TGI) dynamics is extensively used to model the longitudinal response of tumor size in response to drug treatment in preclinical animal models or patients. Pharmacometric TGI models have increasingly been used to characterize drug-exposure response relationships using semimechanistic parameters related to, for instance, direct treatment effects or resistance to personalize drug treatment.^{1,2} Using TGI models, interindividual variation in tumor growth rate, treatment efficacy, and treatment resistance can be quantified and related to patient-specific characteristics.^{1,3} In recent years, TGI models have been integrated with time-to-event models to predict clinical outcomes, such as overall survival, which allow prediction of clinical outcomes based on the patient-specific tumor growth dynamics parameters.^{4–6}

The use of high-dimensional molecular profiling technologies, including next-generation sequencing, to develop personalized treatment schedules is rapidly developing. In particular in oncology, the use of “omics” technologies to characterize tumor-specific molecular differences to predict variation in treatment response is of great interest.⁷ Although both omics and TGI modeling are of relevance toward personalized treatment strategies, pharmacometric TGI models are not frequently directly applied to high-dimensional covariates. In pharmacometric modeling, stepwise covariate inclusion approaches are still the most commonly used approach to include covariates, which is unsuitable for testing of covariates in a high-dimensional setting.

Current analyses of high-dimensional “omics” datasets predicting treatment response are mostly performed using machine learning (ML) methodologies, such as sparse

regression models, random forests, and deep learning, to obtain predictive signatures of treatment response.^{8–10} The majority of studies with ML approaches are based on either dichotomous survival response or clinical response metrics, such as based on the Response Evaluation Criteria in Solid Tumors (RECIST) system,^{11,12} wherein the observed dynamic tumor disease progression profile is reduced into a limited number of categories. These simplified categorical treatment response metrics lack biological or pharmacological relevance, because factors, such as resistance and direct treatment effects, are merged.¹³

A commonly used ML method is the least absolute shrinkage and selection operator (LASSO), which is a linear regression method with ℓ_1 regularization that can be used for high-dimensional analysis, and results in variable selection.¹⁴ Although ML approaches, such as sparse regression models using the LASSO,^{15–18} have been implemented in pharmacometric modeling, they are computationally expensive due to the combination of nonlinearity and estimation of random effects, which often lead to convergence problems. The implementations of the LASSO involve alternating algorithms, which alternate between estimating the random effects and the LASSO optimization, so although LASSO is rather efficient, iterating through multiple random effect estimation steps can severely reduce computational efficiency. This can lead to long computation times and poor convergence rates, especially in high-dimensional settings.

In this study, we propose a two-step approach combining ML, using LASSO regression, with pharmacometric modeling. We demonstrate our approach using a large dataset consisting of longitudinal tumor growth profiles of patient-derived xenograft (PDX) models treated with a variety of mono- and combination regimens.¹⁹ We develop pharmacometric tumor

growth models quantifying intertumor variation in growth rates, drug effect, and resistance, after which we implement ML-based LASSO models to address the following aims: (1) to predict longitudinal tumor growth profiles based on omics-derived predictors using a multivariate LASSO model; and (2) to identify biological pathways associated with interindividual variation in treatment response or resistance development using a group LASSO regression model (Figure 1).

METHODS

Data

Data from a large scale preclinical study in PDX mice models were used.¹⁹ This dataset consisted of over 4000 PDX experiments, which were derived from a total of 277 patients,

where multiple PDX experiments were derived from the same tumor. The PDX experiments from one tumor were all treated with different anticancer agents as mono treatment or combination treatment, or left untreated (e.g., natural growth experiments). There was a total of 62 unique treatments and for every tumor treated one PDX was left untreated, leading to an incomplete design with multiple PDX experiments per treatment. Tumor volume was measured daily. For each unique tumor, at the start of the treatment, genomic data based on gene copy number variations (CNVs) were obtained, yielding a total of 23,852 CNVs. We included data for 174 unique tumors and 55 unique treatments, corresponding with 3244 tumor-treatment combinations. This selection was based on availability of CNV data and adequate fit (see section below). The analysis was conducted separately for every treatment, so the number of observations differed per analysis, ranging from 17 to 171 observations (Table S1).

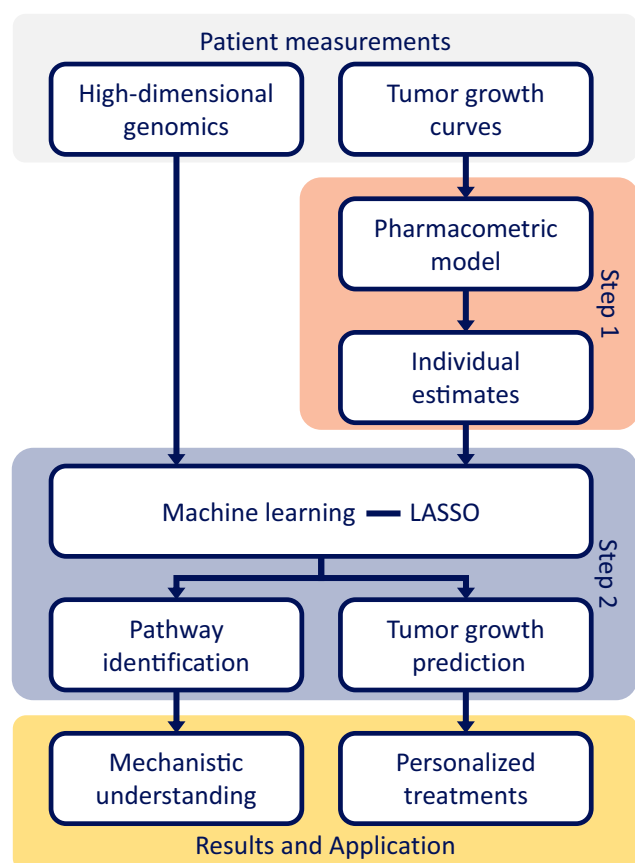


FIGURE 1 A schematic visualization of the proposed two-step approach. First, the tumor growth curves were modeled to obtain tumor growth parameter estimates, second, the individual estimates tumor growth parameter estimates were regressed on copy number variations (genomics) by different least absolute shrinkage and selection operator (LASSO) techniques. The group LASSO was applied to obtain biological pathways. The multivariate LASSO was applied to predict the tumor growth parameter values, which were then inserted into the tumor growth inhibition model equations to obtain predictions of the tumor growth curves

Tumor growth inhibition model

A TGI model was fitted to the longitudinal tumor volume measurements using the nonlinear regression modeling software NONMEM,²⁰ with first order conditional estimation with interaction.⁴ The TGI model captured the longitudinal tumor volume measurements, per PDX, through estimation of three parameters: growth rate (k_g), treatment efficacy (k_d), and time-dependent resistance development (k_r) in an ordinary differential equation (Equation 1).

$$\frac{dV_i(t)}{dt} = k_{g,i} \cdot V_i(t) - k_{d,i} \cdot e^{-k_{r,i} \cdot t} \cdot V_i(t) \quad (1)$$

with tumor volume $V(t)$ at time t and tumor growth model parameters k_g , k_d , and k_r . Random effects with a log-normal distribution, were added to all fixed effect TGI parameters as follows: $k_{g,i} = k_g \cdot \exp(\eta_{k_g,i})$.

To fit the TGI model, we first estimated individual value for k_g separately for every tumor using the untreated PDX data (Equation 2).

$$\frac{dV_i(t)}{dt} = k_{g,i} \cdot V_i(t). \quad (2)$$

The empirical Bayes estimates (EBEs) of k_g were extracted and included as data in the TGI model. EBEs in NONMEM is the estimation of the posterior individual random effects ($\hat{\eta}_i$), based on the empirically obtained prior distribution of η and the individual data, as previously described.²¹ The residual error was modeled with both an additive and proportional error.

We observed that not all tumor growth curves showed time-dependent resistance development (e.g., regrowth), so

both a full TGI model and a reduced model, without a term for resistance, were fitted, effectively allowing k_r to become zero. For every PDX, a likelihood ratio test was conducted to evaluate whether inclusion of k_r added significantly to the model fit (at significance level 0.05). A second criterium was added to only select the full model if the k_r was estimated to be smaller than 1.0, because the term $k_d \cdot e^{-k_r \cdot t}$ goes to zero very fast with t for larger k_r , effectively making k_d unidentifiable.

To evaluate the model fit separately for each treatment, we plotted the conditional weighted residuals per treatment, which represent the goodness of fit for the TGI models.²² Due to the large number of tumor growth curves, treatments with curves with a bad model fit were removed from further analysis. The EBEs of k_d and k_r were extracted for the treatments with good model fit.

Tumor growth profile prediction

The multivariate outcome the log-transformed k_g , log-transformed k_d , and k_r was regressed on the genomic CNV data within every treatment using a multivariate LASSO (Equation 3).²³ The multivariate lasso, similarly to the standard LASSO, minimized the loss function to estimate the linear parameters β is mainly due to outcome Y and parameter β , which are, in this case, both matrices containing a column for every outcome. The penalty term is the root of the summed square error over the vector β_j .

$$\hat{\beta}_{\text{MVlasso}} = \arg \min_{\beta} \left\{ \|Y - X\beta\|_2^2 + \lambda \sum_{j=1}^p \|\beta_j\|_2 \right\} \quad (3)$$

The LASSO hyperparameter λ , which determines the size and number of non-zero parameters, was chosen through 10-fold cross-validation, to identify the λ , which minimized the prediction error in terms of mean squared error. This minimizing λ differed per treatment. The treatments where the minimizing λ did not outperform the null model, which estimated no non-zero coefficients for the CNVs, were removed from further analysis, both in prediction of the tumor growth curves and the pathway selection. In a second analysis, only the log-transformed k_g , log-transformed k_d were regressed on the CNV data. Prediction errors were evaluated both on the scale of the predicted parameter values and on the scale of the predicted tumor growth curves.

The individual TGI parameter values predicted from the LASSO were extracted. The ordinary differential equation (Equation 1) was solved for these predicted parameter values to bring the predictions back on the longitudinal tumor volume scale. For robustness, the cross-validation step was repeated 20 times over different cross-validation

splits and the predicted curves were averaged over the 20 repetitions.

A measure of prediction error was defined on tumor curve scale through comparing the curves from the estimated parameters from the TGI model to the curves with the predicted parameters from LASSO. The prediction error was defined as the absolute fraction of the area between the predicted and the estimated curves (ABC) over the area under the estimated curve (AUC), called the scaled ABC (sABC; Equation 4).

$$\text{sABC}_i(\tau) = \frac{\int_0^\tau |\tilde{V}_i(t) - \hat{V}_i(t)| dt}{\int_0^\tau \tilde{V}_i(t) dt} \quad (4)$$

for individual i with volume $\tilde{V}_i(t)$ estimated from the TGI model fit (IPRED) and volume $\hat{V}_i(t)$ predicted from the multivariate LASSO. The area is considered until some cutoff τ , which in our study was set to 56 days (2 months). The sABC was used because it is a one-dimensional and interpretable error measure. The sABC metric allowed for the comparison of the two functions produced by the TGI model fit and the LASSO parameter value prediction. The sABC of the LASSO with CNVs was compared with the sABC of the null model, to see whether the CNVs added predictive power.

Pathway selection

To gain biological insight gained beyond selection of individual genes contributing to the predictive performance of treatment efficacy and time-dependent resistance development, the log-transformed k_d and k_r were separately regressed on the CNVs through pathway analysis using overlapping group LASSO.^{24,25} The overlapping grouped LASSO uses a combination of the LASSO and the ℓ_2 norm, a square root of the sum of squares of the coefficients, which is also used for RIDGE regression,²⁶ to select variables on a group level (Equation 5). Each of the G groups contain a set of indices $\mathcal{I}_g$, including all parameter indices of the β s in group g . The size of the group is denoted as $|\mathcal{I}_g|$, which is used to scale the penalty to account for the different group sizes.

$$\hat{\beta}_{\text{group lasso}} = \arg \min_{\beta} \left\{ \|y - \beta X\|_2^2 + \lambda \sum_{g=1}^G \sqrt{|\mathcal{I}_g|} \|\beta_{\mathcal{I}_g}\|_2 \right\} \quad (5)$$

The groups were defined as the pathways from the WikiPathways ontology, which contains a comprehensive overview of biological pathways and processes.^{27–29} A total of 5998 CNVs was grouped to one or more pathways.

Again, 10-fold cross-validation was used to identify the λ , which determined how many pathways were selected. While utilizing a combination of ℓ_1 and ℓ_2 penalties, there is only

one hyperparameter in the group lasso.²⁴ Subsequently, part of the discovered correlations between pathways and treatment response was researched in literature for validation. This analysis was conducted in R³⁰ (version 3.6.3) using the library grpregOverlap (<https://github.com/YaohuiZeng/grpregOverlap>).

Code availability

All scripts and models used for the analysis are available on github (<https://github.com/vanhasseltlab/PDX>).

RESULTS

Tumor growth inhibition model development

The TGI model was fitted to the PDX tumor growth curves, separately for every treatment. For three treatments, no model

was converged, these were left out of the analysis. The model fit was evaluated through the conditional weighted residuals (Figure S1) and the visual inspection of the PDX fits (Figure S2). The visual inspection showed the tumor dynamics for treatment TAS266 were not captured. Combination therapy LFW527 and binimetinib showed skewed residual distributions. The two treatments were discarded for further analysis. The model fit for the other treatments was sufficient.

All individual parameter estimates (EBEs) were extracted from the TGI model (Figure 2a). Figure 2b shows how the values of the parameter estimates affect the curve. The percentage of PDX experiments with non-zero time-dependent resistance development was 12.6%. The TGI model for the chosen treatments showed sufficient fits for the next step parameter values prediction step.

The effect of shrinkage of the individual prediction values (η), often referred to as eta-shrinkage, was evaluated in Figure S2. The fit of the individual growth curves was not influenced by shrinkage. Because the tumor volumes were densely sampled over time, with an average of 0.3 samples

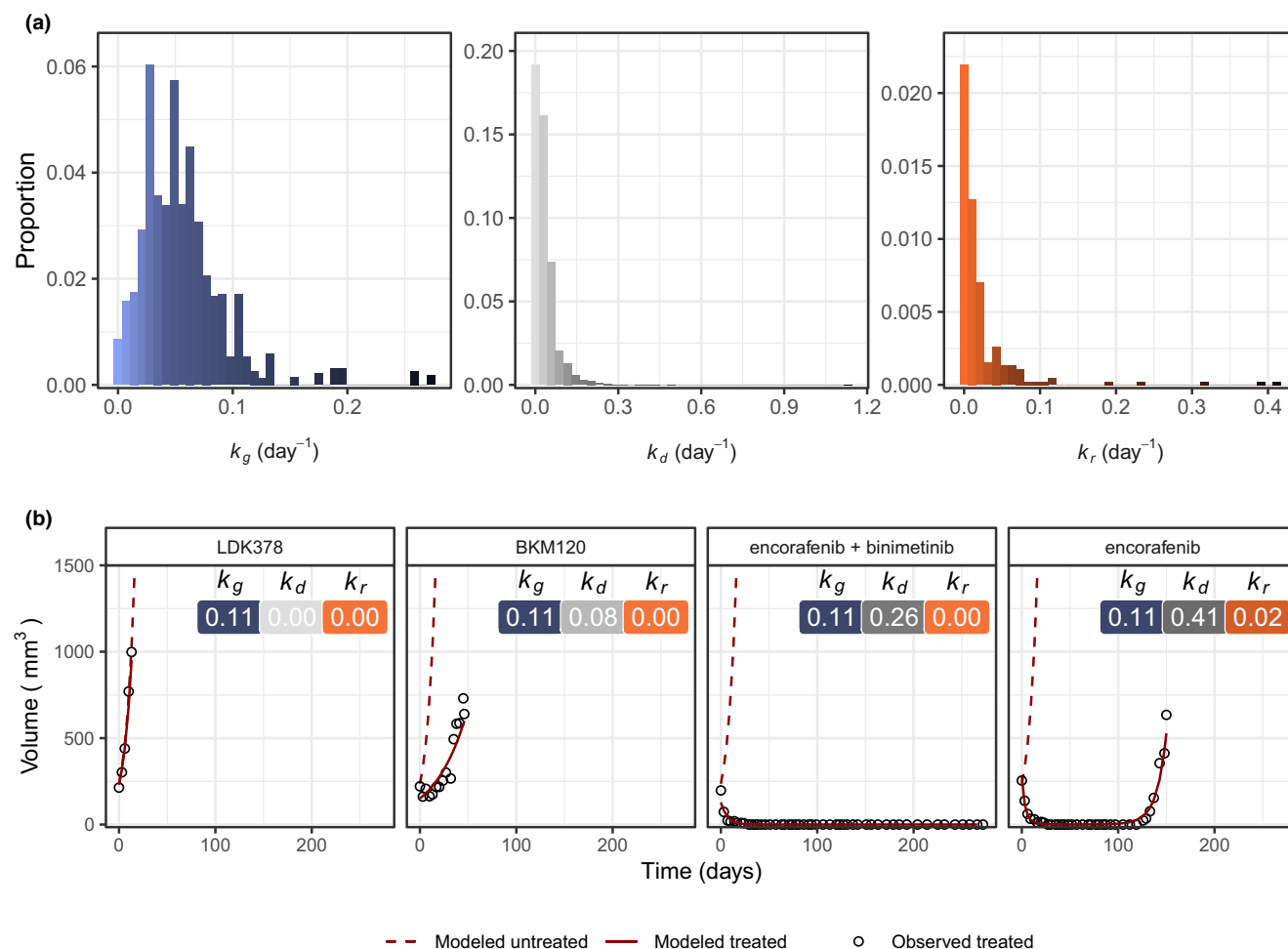


FIGURE 2 Results of the tumor growth inhibition (TGI) model estimation. (a) The distributions of the individual, estimated TGI parameters. (b) Selected tumor growth profiles showing how k_d and k_r vary for different treatments, with from left to right a very ineffective treatment, a slightly effective treatment, a very effective treatment and a very effective treatment with time-dependent resistance development

per day, 50 days follow-up time, and 14 measurements per experiment, we did not expect problems with shrinkage.

Prediction of tumor growth profiles using genomic predictors

The estimated individual TGI parameters k_r , log-transformed k_d , and log-transformed k_g were simultaneously predicted by the multivariate LASSO. The prediction errors for the k_g , k_d , and k_r were calculated as root mean square error (RMSE). Although the variation between the treatments was high (Figure S3a), overall, the RMSE was high. For k_g , the RMSE was 0.035, whereas a mean estimated k_g of 0.0564. For both the k_d and the k_r , the RMSEs 0.044 and 0.049, respectively, were actually higher than the mean estimated k_d (0.033) and the k_r (0.0564), indicating a bad prediction of the tumor growth dynamics from CNVs. A multivariate LASSO with only the log-transformed k_d and log-transformed k_g was also fitted. These two LASSO models were compared based on the prediction error of the log-transformed k_d and log-transformed k_g and the sABC error measure (Figure S3), where the model without predicting the k_r seemed to fit better, especially in the case of combination therapies BYL719 and cetuximab, and BKM120 and LJC049, which was used for consecutive analysis.

Of 52 treatments, 33 treatments were detected with a better prediction than the null model, based on the average MSE over the cross-validation replications. For the other treatments, the predictive ability was not improved by adding CNVs as predictors to the LASSO regression. The log k_g and log k_d were transformed back to their original scale and the parameters were used to solve the ordinary differential equation (Equation 1) from the model.

The predictive performance of the LASSO for predicting the TGI parameter values was evaluated by comparing the curves from the predicted estimates to the curves from the TGI model fit, because the estimated curves were already shown to fit the data well. The predictions and estimations are functions instead of measures, so the scaled area between the curves was calculated as error. The overall median sABC is 0.456, which can be interpreted as the area between the predicted and estimated curve, is less than half the area below the estimated curve (Figure 3a). The sABC distributions for the different treatments are shown (Figure 3b). A lower sABC shows a lower prediction error. There were 23.6% of the curves that had an sABC below 0.2, so the difference between the curves is less than 20% of the AUC of the estimated curve. The treatment LFA102 has a median sABC of only 0.153, indicating a good prediction. The worst predictions are in the treatment LGH447 with a median sABC of 0.867. Compared with the null model, the LASSO reduced the sABC by a median decrease of 3.8%. This shows low predictive ability of the CNVs to predict tumor growth curves.

Identification of pathways associated with treatment efficacy and resistance

The TGI parameter values of k_d and k_r were regressed on CNVs grouped in pathways using the overlapping group LASSO. The group LASSO selected the pathways with predictive power for the 33 treatments where predictiveness was shown in the curve prediction step. Out of the 472 pathways from WikiPathways,²⁹ 71 different pathways were selected for one or more of 19 different treatments, with a total of 118 detected pathway-treatment response correlations (Figures 4, 5). The pathways were specifically correlated to either treatment efficacy or resistance development. More pathways were identified for k_d than k_r , due to smaller variation in k_r .

For paclitaxel, trastuzumab, encorafenib, and figitumumab, the US Food and Drug Administration (FDA) approved drugs administered as monotherapy, we compared identified pathways with literature reports to evaluate their biological validity. We identified 14 pathways for these 4 drugs, of which 9 could be confirmed in literature (Table 1), where we confirm previously described mechanisms were detected through our method.

DISCUSSION

In order to utilize high-dimensional omics data to further advance treatment response prediction and understanding, we developed a two-step approach combining an ML method with pharmacometric modeling.

We showed how CNVs can contribute to prediction of variability in tumor growth dynamics. This approach establishes a practical framework to enable personalized treatment selection or even dose optimization. Even though we have applied our approach to preclinical PDX data, TGI models have been widely used for modeling of clinical tumor size measurements to which our approach can be applied. Pharmacometric models, including TGI models, are typically based on ordinary differential equation (ODE) models, which is why we have chosen to formulate our model as ODEs and not using an analytical expression. Importantly, the use of a TGI model enables further integration with either clinical outcome prediction models⁴¹ or it can be integrated with pharmacokinetic-pharmacodynamic (PK-PD) models for TGI to refine dosing regimens to optimally suppress tumor growth. We expect this approach can also be implemented for the analysis of clinical tumor growth data.

In this study, we have set a cutoff of 56 days to evaluate the ability to back-predict tumor growth profiles; however, the predictions can also be extrapolated over a longer time-span, depending on the nature of available omics-data or specific disease or treatment characteristics. In terms of this sABC, the CNVs did, however, not show great improvement of predictive

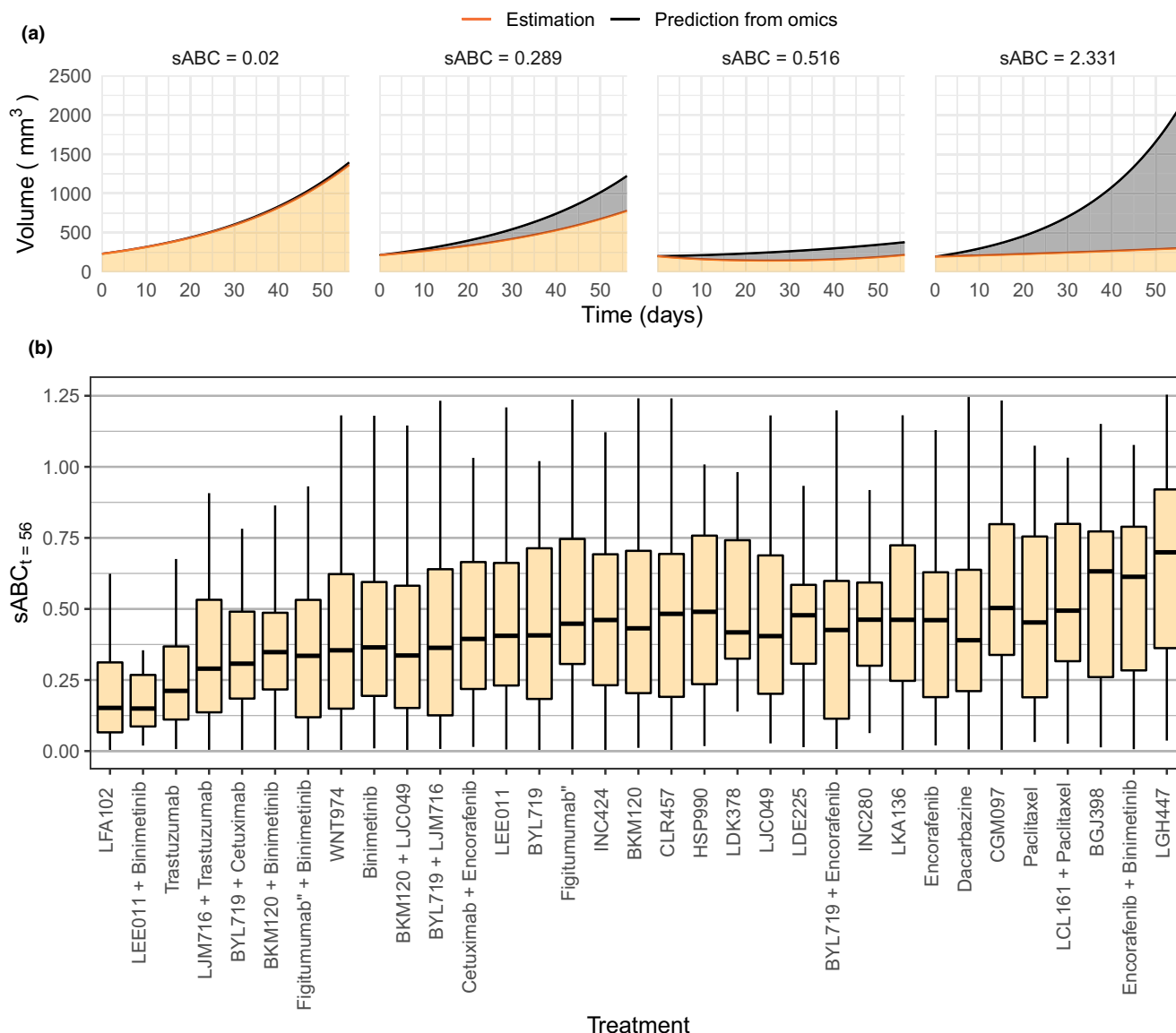


FIGURE 3 Predicted curves from the multivariate least absolute shrinkage and selection operator. (a) Tumor growth curves visualized with the area under the estimated curve (orange) and between the estimated and predicted curves (grey) and the error (in scaled area between the predicted and the estimated curves [sABC]). From left to right show a very good prediction to a very poor prediction. (b) The distributions of the individual patient-derived xenograft sABCs for the different treatments, given by the interquartile range. Outliers are not included in the plot

ability as compared with a null model. This was already visible in the large prediction errors on the tumor growth parameter values. The predictive ability was evaluated instead of the model fit in order to study more generalizable results. A large proportion of variance can often be explained by omics data, but the high-dimensional nature of the data makes it hard or impossible to distinguish between noise and structural differences.

To identify biological factors predictive for either treatment efficacy or resistance development we used a group LASSO, grouping individual gene-associated CNVs to known biological pathways. We have used the WikiPathways ontology for grouping pathways, although other pathway databases can be used in a similar fashion. The pathway group LASSO yields a set of pathways predictive of the outcome

treatment efficacy and time-dependent resistance development. Of 14 identified pathways predictive of treatment efficacy and resistance development, 9 pathways were confirmed by literature search. This is an indication of how omics pathway analysis for dynamic tumor growth responses could be a useful tool for validating pathway associations with factors responsible for treatment response, as well as discovering new correlations with pathways. Such a pathway-oriented approach has been previously proposed, but not in context with TGI or pharmacometric modeling.⁴²

In this study, we have used two versions of LASSO regression for two distinct aims: variable selection and prediction. We selected the use of the LASSO over other ML approaches due to its intrinsic property of variable selection.¹⁴ The



FIGURE 4 Selected pathways obtained by the group least absolute shrinkage and selection operator for the treatment efficacy (k_d = black), time-dependent resistance development (k_r = orange) or both (blue) over the different treatments. The distribution of pathways found for different treatments (top)

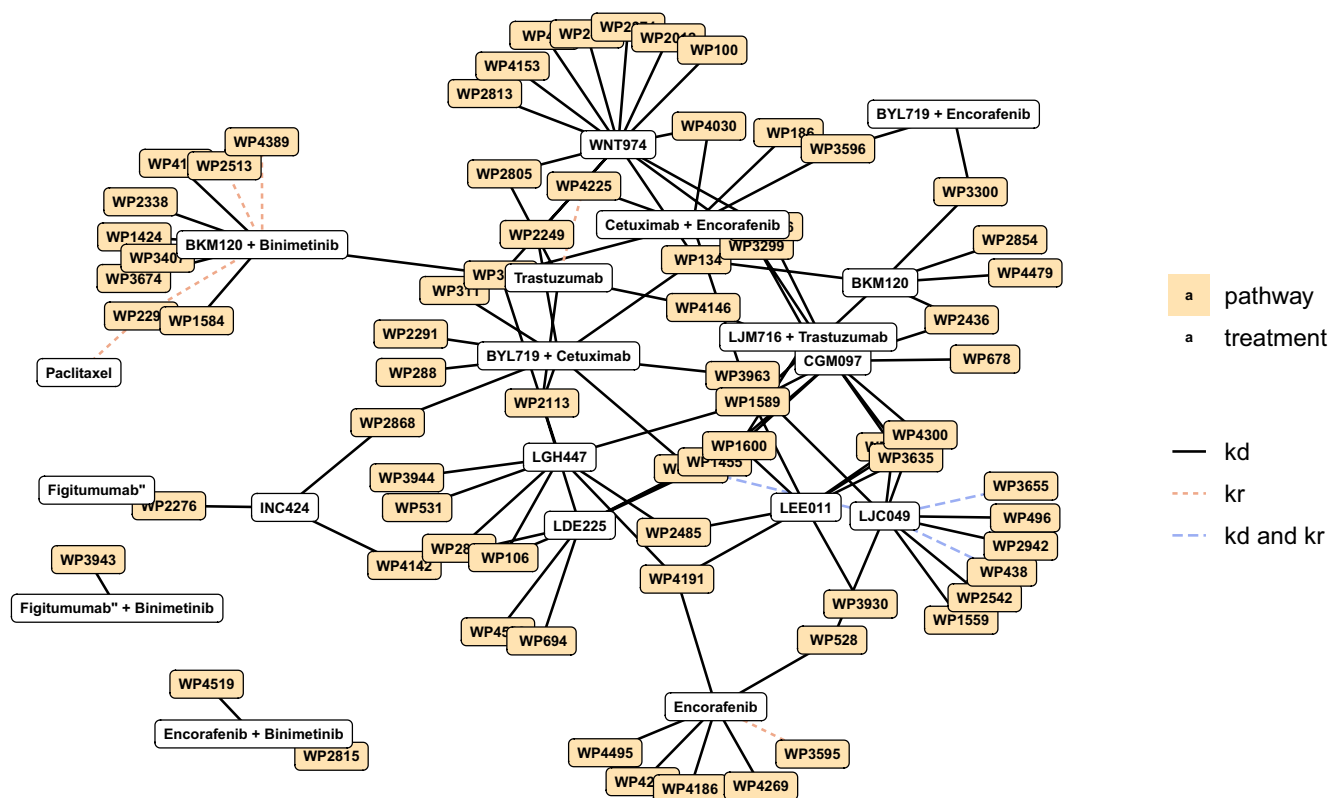


FIGURE 5 Overview of overlapping pathways between the different treatments. The nodes are the treatments (white background) and pathways (orange background) and the edges indicate which tumor growth inhibition parameter links the pathway to the treatment

selection for variables in high-dimensional data is not well accommodated in many algorithms, whereas the LASSO inherently shrinks noise variables to zero. The LASSO can achieve high sensitivity, but it can suffer from low specificity, this, however, is not considered as much of a problem in exploratory analyses.

The use of the group LASSO allows for direct pathway selection based on omics data, which is computationally efficient and interpretable.⁴² The variable selection performance of the LASSO has been investigated previously, and has been shown to perform competitively or comparatively better than other methods.^{43–45}

The multivariate LASSO was used to simultaneously predict the three model parameter values. A limitation of the multivariate LASSO used in this study is that it does not take into account the dependence between the outcome variables, whereas the tumor growth model parameters are expected to be correlated. A second limitation was shown by the comparison between the predictions with and without adding k_r to the multivariate outcome. Prediction of one parameter can restrict the prediction of another parameter. We expect this problem can be overcome by better modeling of the joint and marginal distributions of the multivariate outcome.

The LASSO has been previously implemented in pharmacometric nonlinear mixed effect models.^{16–18} These

direct implementations have the advantage of informing the LASSO directly within the longitudinal modelling. Models with a very high number of variables, however, become computationally hard. To our knowledge, these LASSO implementations have not been successfully applied to very high-dimensional data, where the number of variables (p) was an order of magnitude larger than the number of observations (n), either due to convergence problems, or exploding computation times. The two-step method is more dependent on the fit of the first model and the accuracy of the EBEs. Our method is more feasible in high-dimension, because the steps of the complex longitudinal model estimation and the high-dimensional predictors are separated.

The two-step approach can directly use other ML algorithms besides the LASSO. Algorithms, such as Random Forests and Gradient Boosting, are able to capture non-linearity more easily, and can be used to improve model prediction accuracy. There is still a challenge in modeling multiple outcomes at the same time, such as the k_g , k_d , and k_r in our study, but multivariate outcome modeling extensions have been made in other high-dimensional methods, such as random forests,⁴⁶ which can be also used to predict tumor growth parameter values, as in the second step of our approach.

In summary, we demonstrated how combining machine learning and pharmacometric modeling can be used to gain

TABLE 1 Pathway-treatment correlations found in literature

Treatment	Pathway	Pathway description	Response type	Literature	Relation
Paclitaxel	WP2290	RaA downstream regulated genes	k_r	Ganapathy et al. (2016) ³¹	Paclitaxel is a mitotic inhibitor by stabilizing the microtubule and RaA has been previously shown to disrupt microtubule formation and inducing mitotic catastrophe.
Trastuzumab	WP311	Synthesis and degradation of ketone bodies	k_d	Jobard et al. (2017) ³²	Ketone production was shown to be increased with effective trastuzumab treatment.
	WP4146	Macrophage markers	k_d	Shi et al. (2015) ³³	Trastuzumab interacts with Fcγ receptors on macrophages for the killing of HER2 cancer cells
	WP4225	Pyrimidine metabolism and related diseases	k_d	Ghosh et al. (2009), ³⁴ Liu et al. (2019) ³⁵	The pyrimidine metabolism pathway has been found in previous studies to correlate with drug response to Trastuzumab, based on pathway enrichment analysis in transcriptomics and metabolomics studies
	WP4191	Caloric restriction and aging	k_d	Chappell et al. (2011) ³⁶	There is a connection between Raf/MEK inhibitors and aging
	WP4186	Somatroph axis (GH) and its relationship to dietary restriction and aging	k_d	Chappell et al. (2011) ³⁶	There is a connection between Raf/MEK inhibitors and aging
Encorafenib	WP3595	mir-124 predicted interactions with cell cycle and differentiation	k_r	Ross et al. (2018) ³⁷	Resistance to Encorafenib has been shown to be correlated to cell cycle and differentiation
	WP4269	Ethanol metabolism resulting in production of ROS by CYP2E1	k_d	Friedlander et al. (2019) ³⁸	Ethanol metabolism resulting in production of ROS by CYP2E1 was found to have a connection to the development of melanoma, thus might be related to drug efficacy of encorafenib in melanoma treatment.
	WP4495	IL-10 Anti-inflammatory Signaling Pathway	k_d	Szczepaniak Sloane et al. (2017) ³⁹ , Sumimoto et al. (2006) ⁴⁰	IL-10 has been researched in the context of overexpression of Raf in patients with cancer, showing that IL-10 is an immunosuppressive factor that is decreased by MEK inhibitors

Scientific literature indicating previous findings on the pathways correlated to treatment efficacy and resistance, for the treatments with paclitaxel, trastuzumab, and encorafenib.

Abbreviation: ROS, reactive oxygen species.

pharmacological understanding of factors driving variation in treatment response, and to enable omics-based personalized treatment regimens.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

L.B.Z., K.L.W.D., T.G., J.J.M., P.J.U., and J.G.C.H. wrote the manuscript. J.G.C.H. and L.B.Z. designed the research. L.B.Z., K.L.W.D., M.J., and J.G.C.H. performed the research. L.B.Z., K.L.W.D., and M.J. analyzed the data.

REFERENCES

1. Ribba B, Holford NH, Magni P, et al. A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis. *CPT Pharmacometrics Syst Pharmacol*. 2014;3:e113.
2. Bender BC, Schindler E, Friberg LE. Population pharmacokinetic-pharmacodynamic modelling in oncology: a tool for predicting clinical response. *Br J Clin Pharmacol*. 2013;79:56-71.
3. Rodriguez-Brenes IA, Komarova NL, Wodarz D. Tumor growth dynamics: Insights into evolutionary processes. *Trends Ecol Evol*. 2013;28:597-604.
4. Claret L, Girard P, Hoff PM, et al. Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *J Clin Oncol*. 2009;27:4103-4108.

5. van Hasselt J, Gupta A, Hussein Z, et al. Integrated simulation framework for toxicity, dose intensity, disease progression, and cost effectiveness for castration-resistant prostate cancer treatment with Eribulin. *CPT Pharmacometrics Syst Pharmacol*. 2015;4:374-385.
6. van Hasselt JGC, Gupta A, Hussein Z, et al. Disease progression/clinical outcome model for castration-resistant prostate cancer in patients treated with Eribulin. *CPT Pharmacometrics Syst Pharmacol*. 2015;4:386-395.
7. Shlien A, Malkin D. Copy number variations and cancer. *Genome Med*. 2009;1:62.
8. Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. *Brief Bioinform*. 2019;20:492-503.
9. Nicolò C, Périer C, Prague M, et al. Machine learning and mechanistic modeling for prediction of metastatic relapse in early-stage breast cancer. *JCO Clinical Cancer Informatics*. 2020;4:259-274.
10. Xie G, Dong C, Kong Y, et al. Group lasso regularized deep learning for cancer prognosis from multi-omics and clinical features. *Genes*. 2019;10(3):240.
11. Lathrop K, Kaklamani V. Predictive biomarkers in oncology. In *Predictive Biomarkers in Oncology* (Badve S, Kumar GL eds) 501-511 (Springer International Publishing, Cham, 2019).
12. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45:228-247.
13. Chadeau-Hyam M, Campanella G, Jombart T, et al. Deciphering the complex: methodological overview of statistical models to derive OMICS-based biomarkers. *Environ Mol Mutagen*. 2013;54:542-557.
14. Tibshirani R. Regression shrinkage and selection via the lasso. *Royal Stat Soc*. 1996;58:267-288.
15. Bertrand J, Comets E, Mentre F. Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters. *J Biopharm Stat*. 2008;18:1084-1102.
16. Bertrand J, de Iorio M, Balding DJ. Integrating dynamic mixed-effect modelling and penalized regression to explore genetic association with pharmacokinetics. *Pharmacogenet Genomics*. 2015;25:231-238.
17. Ribbing J, Nyberg J, Caster O, Jonsson EN. The lasso – a novel method for predictive covariate model building in nonlinear mixed effects models. *J Pharmacokinet Pharmacodyn*. 2007;34:485-517.
18. Haem E, Harling K, Ayatollahi TMS, Zare N, Karlsson OM. Adjusted adaptive Lasso for covariate model-building in nonlinear mixed-effect pharmacokinetic models. *J Pharmacokinet Pharmacodyn*. 2017;44:55-66.
19. Gao H, Korn JM, Ferretti S, et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med*. 2015;21:1318-1325.
20. Beal SL, Sheiner LB. *NONMEM User Guide Part VII. NONMEM 7.4 users guides* (ICON plc, Gaithersburg, MD, 1998). <<https://nonmem.iconplc.com/nonmem743/guides>>.
21. Sheiner LB, Rosenberg B, Melmon KL. Modelling of individual pharmacokinetics for computer-aided drug dosage. *Comput Biomed Res*. 1972;5:441-459.
22. Nguyen TH, Mouksassi MS, Holford N, et al. Model evaluation of continuous data pharmacometric models: metrics and graphics. *CPT Pharmacometrics Syst Pharmacol*. 2017;6:87-109.
23. Simon N, Friedman J, Hastie T. A Blockwise Descent Algorithm for Group-penalized Multiresponse and Multinomial Regression; 2013. <<http://arxiv.org/abs/1311.6529>>
24. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *J Royal Stat Soc*. 2006;68:49-67.
25. Jacob L, Obozinski G, Vert J-P. Group lasso with overlap and graph lasso. In *Proceedings of the 26th annual international conference on machine learning* 433-440; 2009.
26. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*. 1970;12:55-67.
27. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44:W90-W97.
28. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14:128.
29. Kutmon M, Riutta A, Nunes N, et al. WikiPathways: capturing the full diversity of pathway knowledge. *Nucleic Acids Res*. 2016;44:D488-D494.
30. R Core Team. R: A Language and Environment for Statistical Computing; 2020. <<https://www.r-project.org/>>.
31. Ganapathy S, Fagman JB, Shen L, et al. Ral A, via activating the mitotic checkpoint, sensitizes cells lacking a functional Nf1 to apoptosis in the absence of protein kinase C. *Oncotarget*. 2016;7:84326-84337.
32. Jobard E, Trédan O, Bachelot T, et al. Longitudinal serum metabolomics evaluation of trastuzumab and everolimus combination as pre-operative treatment for HER-2 positive breast cancer patients. *Oncotarget*. 2017;8:83570-83584.
33. Shi Y, Fan X, Deng H, et al. Trastuzumab triggers phagocytic killing of high HER2 cancer cells in vitro and in vivo by interaction with Fcγ receptors on macrophages. *J Immunol*. 2015;194:4379-4386.
34. Ghosh R, Narasanna A, Gonazalez-Angulo A, Mills G, Arteaga C. Differential signaling by ErbB receptor (HER) dimers: implications for Response to anti-HER2 therapies in breast cancer. *Poster Discussion Abstracts* 69, 705 (American Association for Cancer Research, 2009).
35. Liu W, Wang Q, Chang J. Global metabolomic profiling of trastuzumab resistant gastric cancer cells reveals major metabolic pathways and metabolic signatures based on UHPLC-Q exactive-MS/MS. *RSC Adv*. 2019;9:41192-41208.
36. Chappell WH, Steelman LS, Long JM, et al. Ras/Raf/MEK/ERK and PI3K/PTEN/Akt/mTOR inhibitors: rationale and importance to inhibiting these pathways in human health. *Oncotarget*. 2011;2:135-164.
37. Ross KC, Chin KF, Kim D, et al. Methotrexate sensitizes drug-resistant metastatic melanoma cells to BRAF V600E inhibitors dabrafenib and encorafenib. *Oncotarget*. 2018;9:13324-13336.
38. Friedlander P, Cajulis CB. Melanoma. In *Mount Sinai Expert Guides* 204-213 (Wiley, 2019). <http://doi.org/10.1002/9781119189596.ch18>
39. Szczepaniak Sloane RA, Gopalakrishnan V, Reddy SM, et al. Interaction of molecular alterations with immune response in melanoma. *Cancer*. 2017;123:2130-2142.
40. Sumimoto H, Imabayashi F, Iwata T, Kawakami Y. The BRAF-MAPK signaling pathway is essential for cancer-immune evasion in human melanoma cells. *J Exp Med*. 2006;203:1651-1656.
41. Claret L, Gupta M, Han K, et al. Evaluation of tumor-size response metrics to predict overall survival in western and Chinese patients with first-line metastatic colorectal cancer. *J Clin Oncol*. 2013;31:2110-2114.

42. Silver M, Montana G. Fast identification of biological pathways associated with a quantitative trait using group lasso with overlaps. *Stat Appl Genetics Molecular Biol.* 2012;11:1-43.
43. Lenters V, Vermeulen R, Portengen L. Performance of variable selection methods for assessing the health effects of correlated exposures in case-control studies. *Occup Environ Med.* 2018;75:522-529.
44. Hastie T, Tibshirani R, Tibshirani RJ. Extended Comparisons of Best Subset Selection, Forward Stepwise Selection, and the Lasso. 1-19 (2017). <<http://arxiv.org/abs/1707.08692>>.
45. Zheng S, Liu W. An experimental comparison of gene selection by Lasso and Dantzig selector for cancer classification. *Comput Biol Med.* 2011;41:1033-1040.
46. Segal M, Xiao Y. Multivariate random forests. *WIREs Data Mining Knowledge Discovery.* 2011;1:80-87.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Zwep LB, Duisters KL, Jansen M, et al. Identification of high-dimensional omics-derived predictors for tumor growth dynamics using machine learning and pharmacometric modeling. *CPT Pharmacometrics Syst. Pharmacol.* 2021;10:350–361. <https://doi.org/10.1002/psp4.12603>