Supplementary Methods

Bin-Split binarization method for feature selection

The motivation behind our proposed Bin-Split Symmetrical Binarization approach is to accurately model different effects of high-level and low-level measurements of the lipid species. By using two binary features, our machine learning models can learn two sets of weights for these features. With standard binarization, only one learned binarized feature weight would be obtained, which assumes a strict inverse relationship between high and low levels of each lipid for the task at hand.

In split binarization, specifically, if a numerical lipid species value is greater than the median, it is considered as a high-level case and is represented as (high-level-lipid = 1, low-level-lipid = 0) in the two binary features. Conversely, if the value is lower than the median, it is considered as a low-level case, and the two binary features will be (high-level-lipid = 0, low-level-lipid = 1) in the two binary features. An illustration of split binarization computation is shown in **Figure SM1**.

Test-score computation method in machine learning pipeline

To ensure a thorough and reliable evaluation of our machine learning models, we adopt the widely-used five-fold cross validation method. This method involves dividing the patient data into five equal folds. In each iteration, one fold is designated as the validation set, while the remaining four folds are used for training the model. This process is repeated five times, each time using a different fold as the validation set. Given the metrics, the test score is calculated using the validation set and recorded. After all five iterations, the average of the recorded accuracy scores is taken as the final test score, providing a comprehensive evaluation of the performance of the models. All models were trained using gradient-based optimization algorithms [1].

We use five metrics as the test scores: AUC, accuracy, precision, recall, and specificity. AUC (Area Under the Curve) is a measure of the overall performance of a binary classification model. It represents the area under the receiver operating characteristic (ROC) curve and ranges from 0.5 (random classification) to 1.0 (perfect classification). A higher AUC indicates better model performance. AUC is eqaul to the Concordance index or C-index under the binary classification case. Accuracy is the proportion of correctly classified samples out of the total number of samples. It is calculated as the sum of true positives and true negatives divided by the total number of samples. Precision is the proportion of true positives among all predicted positives. It is calculated as the number of true positives divided by the sum of true positives and false positives. Recall (also known as sensitivity) is the proportion of true positives among all actual positives. It is calculated as the number of true positives divided by the sum of true positives and false negatives. Specificity is the proportion of true negatives among all actual negatives. It is calculated as the number of true negatives divided by the sum of true negatives and false positives. These metrics can be used to evaluate the performance of a medical diagnostic test or model. For example, accuracy can be used to determine how often a test correctly identifies a patient with a certain condition, while precision and recall can be used to evaluate how well the test performs for positive and negative cases, respectively. AUC can provide an overall measure of the test's performance.

Comparation between proposed method and three-signature method

We would like to emphasize the differences between our approach and the three-lipid-signature method presented in [2]. Both methods utilize Logistic regression for metastatic prostate cancer modeling, however, the three-lipid-signature only uses three lipid species (ceramide (d18:1/24:1), sphingosine (d18:2/16:0), and phosphatidylcholine (16:0/16:0)) which may limit the model's capacity. In contrast, our method incorporates a larger number of genetic and lipidomic

features and utilizes elastic-net regularization to implement a more data-driven approach in identifying the most relevant features.

Thus, we post the comparation between proposed method and three-signature method following this procedure. We adopt the survival prediction task over mHSPC patients as the target. To make a fair comparison, we compared the propased machine learning method with standard three-lips-model, three-lids-model with gene features in mHSPC(mutations in ATM, BRCA1, BRCA2,CHEK2) and three-lids-model with bin-split binarized lip features. We randomly select 60% of patients' data for model training and evaluate at the left 40%. For each approch, we compute the true-potive-rate(TPR) and false-positive-rate(FPR) and then draw the ROC curve. We compute the area under the ROC curve (AUC), which is also known as the C-index, as the final evluation metric.

Relative effect ratio computation method in feature analysis

Computing the relative effect ratio is an important step in evaluating the performance of a trained model. The relative effect ratio helps us understand the relative importance of each predictive or protective feature, based on the raw feature weights produced by the model.

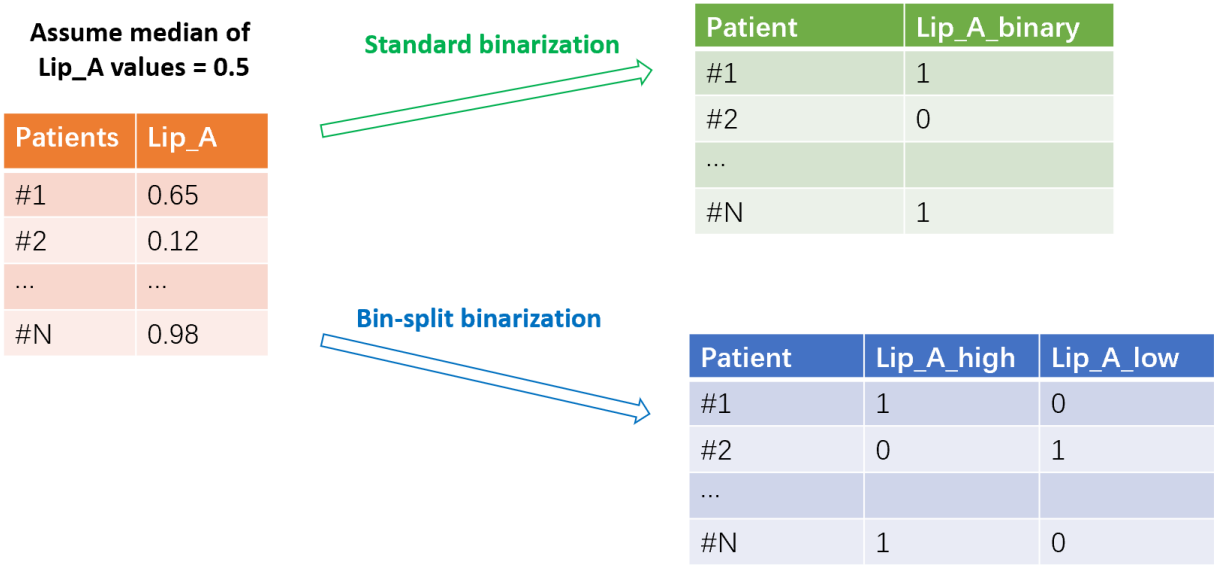
The formula for computing the relative effect ratio involves dividing each positive or negative feature weight by the sum of all positive or negative feature weights. This produces a ratio that measures the relative significance of each feature in the model. By computing the relative effect ratio, we gain insights into the performance of our model. For example, we can identify which features are the most important predictors or protectors, and which features may be less significant or even redundant. This information can be used to improve the model by removing or adjusting redundant features, or by developing new features that capture important patterns in the data.

Predictive probality score methods

To make a clinical prediction model for new patients, based on the feature effects (learned weights) from this study, we initially measure the top-***N*** multi-omics features values $\{x_n\}_{n=1}^{N}$ instead of all features. These top features provide the most positive (if the weight was positive) or negative (if the weight value is negative) effects. Then, we substituted the trained weights $\{w_n\}_{n=1}^{N}$ and the measures of features values $\{x_n\}_{n=1}^{N}$ into a Logistic regression equation, which will yield a probability score. By comparing the prediction score with a pre-set threshold, which we always set to 0.5 we make a final prediction. Figure S1 illustrates this approach for predictive model building.

We also investigate the trade-off effects between the number of features used and the sacrifice in the prediction performance by only picking up the top-***N*** features with largest effects. For each task, we use the corresponding optimal combination of multi-omics features as the candidates, and select top-***N*** features of the candidates with largest effects to build the prediction models. It is clear that we can maintain the high accuracy even only with 50% of features for all tasks. The sparsity effect will allow us to only investigate fewer features for fast and efficient prediction in clinical application.

Figure SM1



References

1. Ruder, S., *An overview of gradient descent optimization algorithms.* arXiv preprint arXiv:1609.04747, 2016.
2. Lin, H.-M., et al., *Aberrations in circulating ceramide levels are associated with poor clinical outcomes across localised and metastatic prostate cancer.* Prostate cancer and prostatic diseases, 2021. **24**(3): p. 860-870.

Supplementary Results

Results for myltiple Machine-Learning approaches

The results revealed that complex non-linear models such as kernel Support Vector Machine (SVM) and Gaussian Process Regression (GPR) were vulnerable to the risks of overfitting, limited data samples, and potential noise perturbations, resulting in worse performance when evaluated using 5-folds cross-validation. Logistic regression with Elastic-Net regularization was observed to be the best-fit model. The performance of the various machine learning methods and feature sets across the four target tasks is presented in **Tables S1- S4.** Names of the genes and lipid species for each table/task are listed in the “Legends” section.

As a result, we selected Logistic regression as our primary model. The inclusion of Elastic-Net regularization was intended to produce a more robust model, sparse feature weights, and highlight the most impactful features during the prediction process. Logistic regression, a linear classification model, calculates a linear combination of features and predicts class probabilities through a logit transformation. Elastic-net regularization, a combination of absolute and squared values of feature weights, helps mitigate overfitting and results in more meaningful regression weights.

Results from Comparing with Three-Lipid-Signatures Method

We evaluated the three-lipid-signature for survival prediction over mHSPC patients, but found that its predictive performance was inferior to our approach. To make a fair comparison and understand the reasons for the difference, we also applied the same data processing approach, Bin-split binarization, to the three-lipid-signature model but still obtained lower performance. We also evaluate the results of three-lipid-signature and gene features(mutations in *ATM, BRCA1, BRCA2,CHEK2*). This indicates the advantage of using multi-omic features and machine

learning methods in our approach. The comparison results are presented in **Table S5**. The ROC curve are plotted in **Figure S1**.

Full list of feature effects analysis

The full lists of all feature effects in each optimal feature sets over four target tasks are shown in **Table S6A-S6H.** We use the raw feature weights to further compute the relative effect ratios as described in *Supplementary Methods*.

Trade-off analysis of feature numbers and model performance

The proposed methods enable us to construct fast prediction for the future patients by only using very limited number of the features, instead of measuring all multi-omic candidates. To investigate the trade-off between using less features and sacrifice of model performance, we plotted the prediction accuracy dropping curve when deceasing the number of features used. The results are shown for all four tasks in **Figure S2-S5** in which we observe that prediction accuracy is maintained decreases when we use 50% or less of all features.

Full list of all species-level lipidomic features

The full name and the corresponding class of all used lipid features are listed in **Table S7**.

Fast model example

We give an exact example on how to build fast prediction model and compute the probability score for future patients on **Figure S6** using three features as an illustrative example.

Table S1:

Features-Combination	Logistic Regression	Logistic Reg Elastic-Net	Linear-kernel SVM	RBF-kernel SVM	RBF-kernel GPR	Matern-kernel GPR
Gene	0.514	0.494	0.563	0.591	0.577	0.563
TG	0.462	0.476	0.421	0.409	0.491	0.406
Cer	0.665	0.665	0.635	0.507	0.62	0.591
DG	0.604	0.576	0.59	0.576	0.506	0.463
Gene_TG	0.462	0.476	0.434	0.409	0.491	0.406
Gene_Cer	0.65	0.665	0.55	0.521	0.62	0.591
Gene_DG	0.59	0.548	0.604	0.605	0.435	0.435
TG_Cer	0.59	0.591	0.521	0.424	0.491	0.562
TG_DG	0.477	0.477	0.521	0.478	0.491	0.433
DG_Cer	0.681	0.681	0.706	0.676	0.676	0.676
Gene_TG_Cer	0.59	0.591	0.534	0.424	0.491	0.549
Gene_TG_DG	0.477	0.477	0.492	0.478	0.491	0.433
Gene_DG_Cer	0.683	0.711	0.681	0.634	0.67	0.67
TG_Cer_DG	0.619	0.634	0.606	0.451	0.491	0.576
Gene_TG_Cer_DG	0.619	0.634	0.606	0.451	0.491	0.576

Table S2:

Features/Methods	Logistic Regression	Logistic Reg Elastic-Net	Linear-kernel SVM	RBF-kernel SVM	RBF-kernel GPR	Matern-kernel GPR
Gene	0.61	0.62	0.62	0.592	0.592	0.606
TG	0.466	0.451	0.493	0.52	0.451	0.466
Cer	0.479	0.507	0.505	0.49	0.549	0.55
DG	0.605	0.59	0.565	0.534	0.508	0.48
Gene_TG	0.466	0.451	0.48	0.52	0.451	0.466
Gene_Cer	0.493	0.508	0.507	0.505	0.534	0.548
Gene_DG	0.639	0.64	0.59	0.549	0.518	0.508
TG_Cer	0.481	0.51	0.481	0.534	0.465	0.425
TG_DG	0.521	0.506	0.536	0.505	0.479	0.479
DG_Cer	0.533	0.519	0.546	0.561	0.531	0.56
Gene_TG_Cer	0.467	0.51	0.48	0.534	0.465	0.425
Gene_TG_DG	0.521	0.506	0.493	0.518	0.479	0.493
Gene_DG_Cer	0.519	0.506	0.561	0.547	0.545	0.519
TG_Cer_DG	0.481	0.509	0.495	0.519	0.465	0.423
Gene_TG_Cer_DG	0.481	0.509	0.508	0.533	0.465	0.423

Table S3:

Features/Methods	Logistic Regression	Logistic Reg Elastic-Net	Linear-kernel SVM	RBF-kernel SVM	RBF-kernel GPR	Matern-kernel GPR
Gene	0.737	0.737	0.737	0.737	0.73	0.73
Cer	0.646	0.66	0.541	0.653	0.668	0.646
Acy	0.667	0.668	0.619	0.653	0.668	0.668
Sph	0.668	0.668	0.668	0.661	0.668	0.668
Gene_Cer	0.709	0.716	0.59	0.681	0.723	0.695
Gene_Acy	0.724	0.703	0.73	0.723	0.668	0.668
Gene_Sph	0.71	0.702	0.732	0.732	0.668	0.668
Acy_Cer	0.646	0.646	0.528	0.647	0.668	0.66
Sph_Cer	0.653	0.66	0.576	0.674	0.668	0.652
Sph_Acy	0.675	0.668	0.598	0.647	0.668	0.668
Gene_Acy_Cer	0.732	0.749	0.576	0.674	0.716	0.716
Gene_Sph_Cer	0.709	0.716	0.576	0.702	0.716	0.702
Gene_Sph_Acy	0.714	0.709	0.702	0.737	0.668	0.668
Cer_Acy_Sph	0.646	0.653	0.534	0.661	0.668	0.618
Gene_Cer_Acy_Sph	0.709	0.709	0.542	0.702	0.702	0.681

Table S4

Features/Methods	Logistic Regression	Logistic Reg Elastic-Net	Linear-kernel SVM	RBF-kernel SVM	RBF-kernel GPR	Matern-kernel GPR
Gene	0.687	0.687	0.687	0.687	0.687	0.687
Cer	0.681	0.687	0.674	0.687	0.687	0.666
Acy	0.687	0.687	0.681	0.687	0.687	0.687
Sph	0.687	0.687	0.687	0.681	0.687	0.687
Gene_Cer	0.701	0.701	0.687	0.681	0.652	0.735
Gene_Acy	0.687	0.687	0.667	0.681	0.687	0.687
Gene_Sph	0.687	0.687	0.687	0.653	0.687	0.687
Acy_Cer	0.681	0.687	0.688	0.687	0.687	0.694
Sph_Cer	0.681	0.687	0.653	0.687	0.687	0.659
Sph_Acy	0.687	0.687	0.674	0.687	0.687	0.687
Gene_Acy_Cer	0.701	0.708	0.674	0.687	0.694	0.715
Gene_Sph_Cer	0.701	0.701	0.694	0.681	0.707	0.715
Gene_Sph_Acy	0.687	0.687	0.674	0.687	0.687	0.687
Cer_Acy_Sph	0.687	0.687	0.633	0.687	0.687	0.687
Gene_Cer_Acy_Sph	0.718	0.722	0.66	0.687	0.687	0.722

Table S5:

mHSPC-Survival	AUC (C-index)
Three Lip Signatures	0.56
Three Lip Signatures with Gene features (mutations in *ATM, BRCA1, BRCA2,CHEK2*)	0.58
Three Lip Signatures with Bin-split binarization	0.52
Multi-omics features with Bin-split binarization (proposed method)	0.71

Table S6A

feature_name	feature_weight	relative_effects_ratio
lip_Cer_Cer(d18:1/14:0)_low	0.723	9.04%
lip_DG_DG(18:0_20:4)_low	0.556	6.95%
lip_Cer_Cer(d18:2/22:0)_high	0.512	6.41%
lip_Cer_Cer(d20:1/22:0)_low	0.465	5.81%
lip_Cer_Cer(d17:1/20:0)_high	0.434	5.43%
lip_DG_DG(18:1_18:2)_low	0.412	5.16%
lip_Cer_Cer(d18:1/26:0)_high	0.412	5.15%
lip_Cer_Cer(d16:1/24:1)_high	0.407	5.10%
lip_Cer_Cer(d19:1/24:1)_high	0.392	4.91%
lip_Cer_Cer(d18:1/22:0)_high	0.384	4.81%
lip_Cer_Cer(m18:0/20:0)_low	0.375	4.69%
lip_Cer_Cer(d20:1/23:0)_low	0.366	4.57%
lip_DG_DG(18:2_18:2)_low	0.359	4.49%
lip_DG_DG(18:1_20:3)_high	0.357	4.47%

lip_Cer_Cer(d16:1/16:0)_low	0.336	4.21%
lip_DG_DG(16:0_16:0)_low	0.319	3.99%
lip_Cer_Cer(m18:1/20:0)_high	0.31	3.87%
lip_Cer_Cer(d17:1/24:1)_high	0.302	3.78%
gene_CHEK2_Mut	0.29	3.62%
lip_DG_DG(14:0_18:2)_high	0.284	3.55%

Table S6B

feature_name	feature_weight	relative_effects_ratio
lip_Cer_Cer(d18:1/14:0)_high	-0.717	9.10%
lip_DG_DG(18:0_20:4)_high	-0.55	6.99%
lip_Cer_Cer(d18:2/22:0)_low	-0.507	6.43%
lip_Cer_Cer(d20:1/22:0)_high	-0.459	5.83%
lip_Cer_Cer(d17:1/20:0)_low	-0.428	5.44%
lip_DG_DG(18:1_18:2)_high	-0.407	5.16%
lip_Cer_Cer(d18:1/26:0)_low	-0.406	5.16%
lip_Cer_Cer(d16:1/24:1)_low	-0.402	5.10%
lip_Cer_Cer(d19:1/24:1)_low	-0.387	4.91%
lip_Cer_Cer(d18:1/22:0)_low	-0.379	4.81%
lip_Cer_Cer(m18:0/20:0)_high	-0.369	4.69%
lip_Cer_Cer(d20:1/23:0)_high	-0.36	4.57%
lip_DG_DG(18:2_18:2)_high	-0.354	4.49%
lip_DG_DG(18:1_20:3)_low	-0.352	4.46%
lip_Cer_Cer(d16:1/16:0)_high	-0.331	4.20%
lip_DG_DG(16:0_16:0)_high	-0.314	3.98%
lip_Cer_Cer(m18:1/20:0)_low	-0.304	3.86%
lip_Cer_Cer(d17:1/24:1)_low	-0.297	3.76%
lip_DG_DG(14:0_18:2)_low	-0.278	3.53%
lip_DG_DG(18:2_22:6)_high	-0.278	3.53%

Table S6C

feature_name	feature_weight	relative_effects_ratio
lip_DG_DG(16:0_22:6)_low	0.880	11.45%
lip_DG_DG(18:2_20:4)_high	0.638	8.31%
lip_DG_DG(18:1_18:2)_high	0.632	8.22%
lip_DG_DG(16:1_18:1)_high	0.548	7.14%
lip_DG_DG(18:0_20:4)_low	0.428	5.57%
lip_DG_DG(18:1_20:3)_high	0.415	5.40%
lip_DG_DG(18:1_18:1)_low	0.396	5.15%
lip_DG_DG(16:0_18:2)_low	0.392	5.10%
lip_DG_DG(14:0_18:2)_high	0.390	5.08%
lip_DG_DG(18:2_22:6)_high	0.369	4.80%
gene_BRCA2_Mut	0.347	4.51%
lip_DG_DG(18:0_22:6)_high	0.342	4.46%
lip_DG_DG(18:1_22:6)_high	0.305	3.97%
lip_DG_DG(16:0_16:1)_low	0.252	3.28%
lip_DG_DG(18:0_18:1)_low	0.244	3.18%
lip_DG_DG(18:1_18:3)_low	0.207	2.69%
lip_DG_DG(18:1_20:4)_low	0.183	2.39%
lip_DG_DG(18:1_20:5)_high	0.182	2.37%
lip_DG_DG(14:0_16:0)_high	0.145	1.89%
lip_DG_DG(16:0_16:0)_high	0.139	1.81%

Table S6D

feature_name	feature_weight	relative_effects_ratio
gene_CHEK2_SNP	-1.127	13.19%
lip_DG_DG(16:0_22:6)_high	-0.852	9.97%
gene_ATM_Mut	-0.727	8.50%
lip_DG_DG(18:2_20:4)_low	-0.610	7.14%
lip_DG_DG(18:1_18:2)_low	-0.604	7.06%
lip_DG_DG(16:1_18:1)_low	-0.520	6.09%
lip_DG_DG(18:0_20:4)_high	-0.400	4.68%

lip_DG_DG(18:1_20:3)_low	-0.387	4.53%
lip_DG_DG(18:1_18:1)_high	-0.367	4.30%
lip_DG_DG(16:0_18:2)_high	-0.363	4.25%
lip_DG_DG(14:0_18:2)_low	-0.362	4.24%
lip_DG_DG(18:2_22:6)_low	-0.341	3.99%
lip_DG_DG(18:0_22:6)_low	-0.314	3.68%
lip_DG_DG(18:1_22:6)_low	-0.277	3.24%
lip_DG_DG(16:0_16:1)_high	-0.224	2.62%
lip_DG_DG(18:0_18:1)_high	-0.216	2.53%
lip_DG_DG(18:1_18:3)_high	-0.179	2.09%
lip_DG_DG(18:1_20:4)_high	-0.155	1.82%
lip_DG_DG(18:1_20:5)_low	-0.154	1.81%

Table S6E

feature_name	feature_weight	relative_effects_ratio
gene_RB1_Del	2.208	15.37%
lip_Cer_Cer(d18:1/17:0)_low	0.699	4.87%
lip_Cer_Cer(d20:1/23:0)_high	0.681	4.74%
lip_Sph_Sph(d16:1)_low	0.592	4.12%
lip_Cer_Cer(d18:1/24:1)_high	0.537	3.74%
lip_Cer_Cer(d17:1/24:1)_high	0.53	3.69%
lip_Cer_Cer(d19:1/24:0)_low	0.527	3.67%
lip_Cer_Cer(d17:1/18:0)_low	0.511	3.56%
lip_Cer_Cer(d20:1/22:0)_high	0.504	3.51%
lip_Cer_Cer(m18:1/18:0)_high	0.5	3.48%
lip_Cer_Cer(d16:1/20:0)_low	0.481	3.35%
lip_Cer_Cer(d18:1/26:0)_low	0.461	3.21%
lip_Sph_Sph(d18:1)_high	0.458	3.19%
gene_AR_Amp	0.442	3.08%
lip_Cer_Cer(m18:1/22:0)_high	0.436	3.04%
lip_Cer_Cer(d18:1/20:0)_high	0.432	3.01%

lip_Cer_Cer(d18:1/21:0)_low	0.401	2.79%
lip_Cer_Cer(d19:1/26:0)_high	0.394	2.74%
lip_Cer_Cer(d18:1/24:0)_low	0.393	2.74%
lip_Cer_Cer(m18:0/20:0)_low	0.376	2.62%

Table S6F

feature_name	feature_weight	relative_effects_ratio
lip_Cer_Cer(d18:1/17:0)_high	-0.741	5.45%
lip_Cer_Cer(d20:1/23:0)_low	-0.723	5.31%
lip_Sph_Sph(d16:1)_high	-0.634	4.66%
lip_Cer_Cer(d18:1/24:1)_low	-0.579	4.26%
lip_Cer_Cer(d17:1/24:1)_low	-0.572	4.20%
lip_Cer_Cer(d19:1/24:0)_high	-0.569	4.18%
lip_Cer_Cer(d17:1/18:0)_high	-0.553	4.07%
lip_Cer_Cer(d20:1/22:0)_low	-0.546	4.01%
lip_Cer_Cer(m18:1/18:0)_low	-0.541	3.98%
lip_Cer_Cer(d16:1/20:0)_high	-0.523	3.84%
lip_Cer_Cer(d18:1/26:0)_high	-0.503	3.70%
lip_Sph_Sph(d18:1)_low	-0.500	3.68%
gene_TP53_Mut	-0.479	3.52%
lip_Cer_Cer(m18:1/22:0)_low	-0.478	3.52%
lip_Cer_Cer(d18:1/20:0)_low	-0.474	3.48%
lip_Cer_Cer(d18:1/21:0)_high	-0.443	3.25%
lip_Cer_Cer(d19:1/26:0)_low	-0.436	3.21%
lip_Cer_Cer(d18:1/24:0)_high	-0.435	3.20%
lip_Cer_Cer(m18:0/20:0)_high	-0.417	3.07%
lip_Cer_Cer(d16:1/22:0)_high	-0.414	3.04%

Table S6G

feature_name	feature_weight	relative_effects_ratio
lip_Cer_Cer(d19:1/26:0)_low	0.686	6.54%
lip_Cer_Cer(d19:1/24:0)_high	0.543	5.18%
lip_Cer_Cer(d18:1/24:0)_high	0.523	4.99%
lip_Cer_Cer(d18:2/24:0)_low	0.521	4.97%
lip_Cer_Cer(d18:1/17:0)_high	0.492	4.70%
lip_Cer_Cer(d20:1/23:0)_low	0.488	4.65%
lip_Cer_Cer(d18:1/22:0)_high	0.453	4.32%
lip_Cer_Cer(d18:1/24:1)_low	0.440	4.19%
lip_Cer_Cer(d19:1/18:0)_high	0.420	4.01%
lip_Cer_Cer(d19:1/20:0)_low	0.404	3.86%
lip_Cer_Cer(d18:1/14:0)_low	0.398	3.79%
lip_Cer_Cer(d18:2/24:1)_high	0.393	3.74%
lip_Cer_Cer(d16:1/24:0)_high	0.355	3.39%
lip_Cer_Cer(d16:1/23:0)_high	0.347	3.31%
lip_Cer_Cer(m18:0/22:0)_high	0.331	3.16%
lip_Cer_Cer(d17:1/23:0)_low	0.329	3.13%
lip_Cer_Cer(d18:1/26:0)_high	0.327	3.12%
lip_Cer_Cer(m18:1/24:1)_high	0.301	2.87%
lip_Cer_Cer(d18:1/19:0)_low	0.299	2.86%
lip_Cer_Cer(d16:1/18:0)_high	0.281	2.68%

Table S6H

Feature_name	Feature_weight	Relative_effects_ratio
gene_AR_Amp	-0.940	7.49%
gene_RB1_Del	-0.850	6.78%
lip_Cer_Cer(d19:1/26:0)_high	-0.698	5.56%
lip_Cer_Cer(d19:1/24:0)_low	-0.555	4.42%
lip_Cer_Cer(d18:1/24:0)_low	-0.535	4.26%

lip_Cer_Cer(d18:2/24:0)_high	-0.533	4.25%
lip_Cer_Cer(d18:1/17:0)_low	-0.505	4.02%
lip_Cer_Cer(d20:1/23:0)_high	-0.500	3.98%
gene_TP53_Mut	-0.485	3.86%
lip_Cer_Cer(d18:1/22:0)_low	-0.465	3.71%
lip_Cer_Cer(d18:1/24:1)_high	-0.452	3.60%
lip_Cer_Cer(d19:1/18:0)_low	-0.432	3.44%
lip_Cer_Cer(d19:1/20:0)_high	-0.417	3.32%
lip_Cer_Cer(d18:1/14:0)_high	-0.410	3.27%
lip_Cer_Cer(d18:2/24:1)_low	-0.405	3.23%
lip_Cer_Cer(d16:1/24:0)_low	-0.368	2.93%
lip_Cer_Cer(d16:1/23:0)_low	-0.359	2.86%
lip_Cer_Cer(m18:0/22:0)_low	-0.344	2.74%
lip_Cer_Cer(d17:1/23:0)_high	-0.341	2.72%
lip_Cer_Cer(d18:1/26:0)_low	-0.339	2.70%

Table S7: The name and type of all lipidomic species features included in modeling and analysis.

Index	Lipid name	Lipid type
1	Sph(d16:1)	Sphingosine(Sph)
2	Sph(d18:1)	Sphingosine (Sph)
3	Sph(d18:2)	Sphingosine (Sph)
4	Cer(d16:1/16:0)	Ceramide(Cer)
5	Cer(d16:1/18:0)	Ceramide(Cer)
6	Cer(d16:1/20:0)	Ceramide(Cer)
7	Cer(d16:1/22:0)	Ceramide(Cer)
8	Cer(d16:1/23:0)	Ceramide(Cer)
9	Cer(d16:1/24:0)	Ceramide(Cer)
10	Cer(d16:1/24:1)	Ceramide(Cer)
11	Cer(d17:1/16:0)	Ceramide(Cer)
12	Cer(d17:1/18:0)	Ceramide(Cer)
13	Cer(d17:1/20:0)	Ceramide(Cer)

14	Cer(d17:1/22:0)	Ceramide(Cer)
15	Cer(d17:1/23:0)	Ceramide(Cer)
16	Cer(d17:1/24:0)	Ceramide(Cer)
17	Cer(d17:1/24:1)	Ceramide(Cer)
18	Cer(d18:1/14:0)	Ceramide(Cer)
19	Cer(d18:1/16:0)	Ceramide(Cer)
20	Cer(d18:1/17:0)	Ceramide(Cer)
21	Cer(d18:1/18:0)	Ceramide(Cer)
22	Cer(d18:1/19:0)	Ceramide(Cer)
23	Cer(d18:1/20:0)	Ceramide(Cer)
24	Cer(d18:1/21:0)	Ceramide(Cer)
25	Cer(d18:1/22:0)	Ceramide(Cer)
26	Cer(d18:1/23:0)	Ceramide(Cer)
27	Cer(d18:1/24:0)	Ceramide(Cer)
28	Cer(d18:1/24:1)	Ceramide(Cer)
29	Cer(d18:1/26:0)	Ceramide(Cer)
30	Cer(d18:2/14:0)	Ceramide(Cer)
31	Cer(d18:2/16:0)	Ceramide(Cer)
32	Cer(d18:2/17:0)	Ceramide(Cer)
33	Cer(d18:2/18:0)	Ceramide(Cer)
34	Cer(d18:2/20:0)	Ceramide(Cer)
35	Cer(d18:2/21:0)	Ceramide(Cer)
36	Cer(d18:2/22:0)	Ceramide(Cer)
37	Cer(d18:2/23:0)	Ceramide(Cer)
38	Cer(d18:2/24:0)	Ceramide(Cer)
39	Cer(d18:2/24:1)	Ceramide(Cer)
40	Cer(d18:2/26:0)	Ceramide(Cer)
41	Cer(d19:1/16:0)	Ceramide(Cer)
42	Cer(d19:1/18:0)	Ceramide(Cer)
43	Cer(d19:1/20:0)	Ceramide(Cer)
44	Cer(d19:1/22:0)	Ceramide(Cer)
45	Cer(d19:1/23:0)	Ceramide(Cer)
46	Cer(d19:1/24:0)	Ceramide(Cer)

47	Cer(d19:1/24:1)	Ceramide(Cer)
48	Cer(d19:1/26:0)	Ceramide(Cer)
49	Cer(d20:1/22:0)	Ceramide(Cer)
50	Cer(d20:1/23:0)	Ceramide(Cer)
51	Cer(d20:1/24:0)	Ceramide(Cer)
52	Cer(d20:1/24:1)	Ceramide(Cer)
53	Cer(d20:1/26:0)	Ceramide(Cer)
54	Cer(m18:0/20:0)	Ceramide(Cer)
55	Cer(m18:0/22:0)	Ceramide(Cer)
56	Cer(m18:0/23:0)	Ceramide(Cer)
57	Cer(m18:0/24:0)	Ceramide(Cer)
58	Cer(m18:0/24:1)	Ceramide(Cer)
59	Cer(m18:1/18:0)	Ceramide(Cer)
60	Cer(m18:1/20:0)	Ceramide(Cer)
61	Cer(m18:1/22:0)	Ceramide(Cer)
62	Cer(m18:1/23:0)	Ceramide(Cer)
63	Cer(m18:1/24:0)	Ceramide(Cer)
64	Cer(m18:1/24:1)	Ceramide(Cer)
65	AC(12:0)	Acylcarnitine(AC)
66	AC(13:0)	Acylcarnitine(AC)
67	AC(14:0)	Acylcarnitine(AC)
68	AC(14:1)	Acylcarnitine(AC)
69	AC(14:2)	Acylcarnitine(AC)
70	AC(15:0) (a)	Acylcarnitine(AC)
71	AC(15:0) (b)	Acylcarnitine(AC)
72	AC(16:0)	Acylcarnitine(AC)
73	AC(16:1)	Acylcarnitine(AC)
74	AC(17:0) (a)	Acylcarnitine(AC)
75	AC(17:0) (b)	Acylcarnitine(AC)
76	AC(18:0)	Acylcarnitine(AC)
77	AC(18:1)	Acylcarnitine(AC)
78	AC(18:2)	Acylcarnitine(AC)
79	DG(14:0_16:0)	Diacylglycerol (DG)

80	DG(16:0_16:0)	Diacylglycerol (DG)
81	DG(16:0_16:1)	Diacylglycerol (DG)
82	DG(14:0_18:2)	Diacylglycerol (DG)
83	DG(16:0_18:1)	Diacylglycerol (DG)
84	DG(16:1_18:1)	Diacylglycerol (DG)
85	DG(16:0_18:2)	Diacylglycerol (DG)
86	DG(18:0_18:1)	Diacylglycerol (DG)
87	DG(18:1_18:1)	Diacylglycerol (DG)
88	DG(18:0_18:2)	Diacylglycerol (DG)
89	DG(18:1_18:2)	Diacylglycerol (DG)
90	DG(18:2_18:2)	Diacylglycerol (DG)
91	DG(18:1_18:3)	Diacylglycerol (DG)
92	DG(16:0_20:4)	Diacylglycerol (DG)
93	DG(18:1_20:3)	Diacylglycerol (DG)
94	DG(18:0_20:4)	Diacylglycerol (DG)
95	DG(18:1_20:4)	Diacylglycerol (DG)
96	DG(16:0_22:5)	Diacylglycerol (DG)
97	DG(18:2_20:4)	Diacylglycerol (DG)
98	DG(16:0_22:6)	Diacylglycerol (DG)
99	DG(18:1_20:5)	Diacylglycerol (DG)
100	DG(18:1_22:5)	Diacylglycerol (DG)
101	DG(18:0_22:6)	Diacylglycerol (DG)
102	DG(18:1_22:6)	Diacylglycerol (DG)
103	DG(18:2_22:6)	Diacylglycerol (DG)
104	TG(48:0) [SIM]	Triacylglycerol (TG)
105	TG(48:1) [SIM]	Triacylglycerol (TG)
106	TG(48:2) [SIM]	Triacylglycerol (TG)
107	TG(48:3) [SIM]	Triacylglycerol (TG)
108	TG(49:1) [SIM]	Triacylglycerol (TG)
109	TG(50:0) [SIM]	Triacylglycerol (TG)
110	TG(50:1) [SIM]	Triacylglycerol (TG)
111	TG(50:2) [SIM]	Triacylglycerol (TG)
112	TG(50:3) [SIM]	Triacylglycerol (TG)

113	TG(50:4) [SIM]	Triacylglycerol (TG)
114	TG(51:0) [SIM]	Triacylglycerol (TG)
115	TG(51:1) [SIM]	Triacylglycerol (TG)
116	TG(51:2) [SIM]	Triacylglycerol (TG)
117	TG(52:1) [SIM]	Triacylglycerol (TG)
118	TG(52:2) [SIM]	Triacylglycerol (TG)
119	TG(52:3) [SIM]	Triacylglycerol (TG)
120	TG(52:4) [SIM]	Triacylglycerol (TG)
121	TG(52:5) [SIM]	Triacylglycerol (TG)
122	TG(53:2) [SIM]	Triacylglycerol (TG)
123	TG(54:0) [SIM]	Triacylglycerol (TG)
124	TG(54:1) [SIM]	Triacylglycerol (TG)
125	TG(54:2) [SIM]	Triacylglycerol (TG)
126	TG(54:3) [SIM]	Triacylglycerol (TG)
127	TG(54:4) [SIM]	Triacylglycerol (TG)
128	TG(54:5) [SIM]	Triacylglycerol (TG)
129	TG(54:6) [SIM]	Triacylglycerol (TG)
130	TG(54:7) [SIM]	Triacylglycerol (TG)
131	TG(56:6) [SIM]	Triacylglycerol (TG)
132	TG(56:7) [SIM]	Triacylglycerol (TG)
133	TG(56:8) [SIM]	Triacylglycerol (TG)
134	TG(56:9) [SIM]	Triacylglycerol (TG)
135	TG(58:10) [SIM]	Triacylglycerol (TG)
136	TG(58:8) [SIM]	Triacylglycerol (TG)
137	TG(58:9) [SIM]	Triacylglycerol (TG)
138	TG(48:0) [NL-16:0]	Triacylglycerol (TG)
139	TG(48:0) [NL-18:0]	Triacylglycerol (TG)
140	TG(48:1) [NL-16:1]	Triacylglycerol (TG)
141	TG(48:1) [NL-18:1]	Triacylglycerol (TG)
142	TG(48:2) [NL-14:0]	Triacylglycerol (TG)
143	TG(48:2) [NL-14:1]	Triacylglycerol (TG)
144	TG(48:2) [NL-16:1]	Triacylglycerol (TG)
145	TG(48:2) [NL-18:2]	Triacylglycerol (TG)

146	TG(48:3) [NL-14:0]	Triacylglycerol (TG)
147	TG(48:3) [NL-16:1]	Triacylglycerol (TG)
148	TG(48:3) [NL-18:3]	Triacylglycerol (TG)
149	TG(49:1) [NL-16:1]	Triacylglycerol (TG)
150	TG(49:1) [NL-17:1]	Triacylglycerol (TG)
151	TG(50:0) [NL-18:0]	Triacylglycerol (TG)
152	TG(50:1) [NL-14:0]	Triacylglycerol (TG)
153	TG(50:1) [NL-16:0]	Triacylglycerol (TG)
154	TG(50:1) [NL-18:1]	Triacylglycerol (TG)
155	TG(50:2) [NL-14:0]	Triacylglycerol (TG)
156	TG(50:2) [NL-16:1]	Triacylglycerol (TG)
157	TG(50:2) [NL-18:1]	Triacylglycerol (TG)
158	TG(50:2) [NL-18:2]	Triacylglycerol (TG)
159	TG(50:3) [NL-14:0]	Triacylglycerol (TG)
160	TG(50:3) [NL-14:1]	Triacylglycerol (TG)
161	TG(50:3) [NL-16:1]	Triacylglycerol (TG)
162	TG(50:3) [NL-18:2]	Triacylglycerol (TG)
163	TG(50:3) [NL-18:3]	Triacylglycerol (TG)
164	TG(50:4) [NL-14:0]	Triacylglycerol (TG)
165	TG(50:4) [NL-18:3]	Triacylglycerol (TG)
166	TG(50:4) [NL-20:4]	Triacylglycerol (TG)
167	TG(51:0) [NL-16:0]	Triacylglycerol (TG)
168	TG(51:1) [NL-17:0]	Triacylglycerol (TG)
169	TG(51:2) [NL-15:0]	Triacylglycerol (TG)
170	TG(51:2) [NL-17:0]	Triacylglycerol (TG)
171	TG(51:2) [NL-17:1]	Triacylglycerol (TG)
172	TG(52:1) [NL-18:0]	Triacylglycerol (TG)
173	TG(52:1) [NL-18:1]	Triacylglycerol (TG)
174	TG(52:2) [NL-16:0]	Triacylglycerol (TG)
175	TG(52:2) [NL-18:2]	Triacylglycerol (TG)
176	TG(52:3) [NL-16:1]	Triacylglycerol (TG)
177	TG(52:3) [NL-18:2]	Triacylglycerol (TG)
178	TG(52:4) [NL-16:1]	Triacylglycerol (TG)

179	TG(52:4) [NL-18:2]	Triacylglycerol (TG)
180	TG(52:4) [NL-18:3]	Triacylglycerol (TG)
181	TG(52:5) [NL-18:3]	Triacylglycerol (TG)
182	TG(52:5) [NL-20:4]	Triacylglycerol (TG)
183	TG(52:5) [NL-20:5]	Triacylglycerol (TG)
184	TG(53:2) [NL-17:1]	Triacylglycerol (TG)
185	TG(53:2) [NL-18:1]	Triacylglycerol (TG)
186	TG(54:0) [NL-18:0]	Triacylglycerol (TG)
187	TG(54:1) [NL-18:1]	Triacylglycerol (TG)
188	TG(54:2) [NL-18:0]	Triacylglycerol (TG)
189	TG(54:2) [NL-20:1]	Triacylglycerol (TG)
190	TG(54:3) [NL-18:1]	Triacylglycerol (TG)
191	TG(54:3) [NL-18:2]	Triacylglycerol (TG)
192	TG(54:4) [NL-18:2]	Triacylglycerol (TG)
193	TG(54:4) [NL-20:3]	Triacylglycerol (TG)
194	TG(54:5) [NL-18:3]	Triacylglycerol (TG)
195	TG(54:5) [NL-20:4]	Triacylglycerol (TG)
196	TG(54:6) [NL-18:3]	Triacylglycerol (TG)
197	TG(54:6) [NL-20:4]	Triacylglycerol (TG)
198	TG(54:6) [NL-20:5]	Triacylglycerol (TG)
199	TG(54:6) [NL-22:6]	Triacylglycerol (TG)
200	TG(54:7) [NL-20:5]	Triacylglycerol (TG)
201	TG(54:7) [NL-22:6]	Triacylglycerol (TG)
202	TG(56:6) [NL-20:4]	Triacylglycerol (TG)
203	TG(56:6) [NL-22:5](a)	Triacylglycerol (TG)
204	TG(56:6) [NL-22:5](b)	Triacylglycerol (TG)
205	TG(56:7) [NL-20:4]	Triacylglycerol (TG)
206	TG(56:7) [NL-20:5]	Triacylglycerol (TG)
207	TG(56:7) [NL-22:5](a)	Triacylglycerol (TG)

208	TG(56:7) [NL-22:5](b)	Triacylglycerol (TG)
209	TG(56:7) [NL-22:6]	Triacylglycerol (TG)
210	TG(56:8) [NL-20:4]	Triacylglycerol (TG)
211	TG(56:8) [NL-20:5]	Triacylglycerol (TG)
212	TG(56:8) [NL-22:6]	Triacylglycerol (TG)
213	TG(56:9) [NL-22:6]	Triacylglycerol (TG)
214	TG(58:10) [NL-22:6]	Triacylglycerol (TG)
215	TG(58:8) [NL-22:6]	Triacylglycerol (TG)
216	TG(58:9) [NL-22:6]	Triacylglycerol (TG)
217	TG(O-50:1) [SIM]	Triacylglycerol (TG)
218	TG(O-50:2) [SIM]	Triacylglycerol (TG)
219	TG(O-50:3) [SIM]	Triacylglycerol (TG)
220	TG(O-52:0) [SIM]	Triacylglycerol (TG)
221	TG(O-52:1) [SIM]	Triacylglycerol (TG)
222	TG(O-52:2) [SIM]	Triacylglycerol (TG)
223	TG(O-54:2) [SIM]	Triacylglycerol (TG)
224	TG(O-54:3) [SIM]	Triacylglycerol (TG)
225	TG(O-54:4) [SIM]	Triacylglycerol (TG)
226	TG(O-50:1) [NL-15:0]	Triacylglycerol (TG)
227	TG(O-50:1) [NL-16:0]	Triacylglycerol (TG)
228	TG(O-50:1) [NL-17:1]	Triacylglycerol (TG)
229	TG(O-50:1) [NL-18:1]	Triacylglycerol (TG)
230	TG(O-50:2) [NL-16:1]	Triacylglycerol (TG)
231	TG(O-50:2) [NL-18:1]	Triacylglycerol (TG)
232	TG(O-50:2) [NL-18:2]	Triacylglycerol (TG)
233	TG(O-50:3) [NL-18:2]	Triacylglycerol (TG)
234	TG(O-52:0) [NL-16:0]	Triacylglycerol (TG)
235	TG(O-52:1) [NL-16:0]	Triacylglycerol (TG)
236	TG(O-52:1) [NL-18:1]	Triacylglycerol (TG)
237	TG(O-52:2) [NL-16:0]	Triacylglycerol (TG)
238	TG(O-52:2) [NL-17:1]	Triacylglycerol (TG)
239	TG(O-52:2) [NL-18:1]	Triacylglycerol (TG)

240	TG(O-54:2) [NL-17:1]	Triacylglycerol (TG)
241	TG(O-54:2) [NL-18:1]	Triacylglycerol (TG)
242	TG(O-54:3) [NL-17:1]	Triacylglycerol (TG)
243	TG(O-54:3) [NL-18:1]	Triacylglycerol (TG)
244	TG(O-54:4) [NL-17:1]	Triacylglycerol (TG)
245	TG(O-54:4) [NL-18:2]	Triacylglycerol (TG)

Figure S1

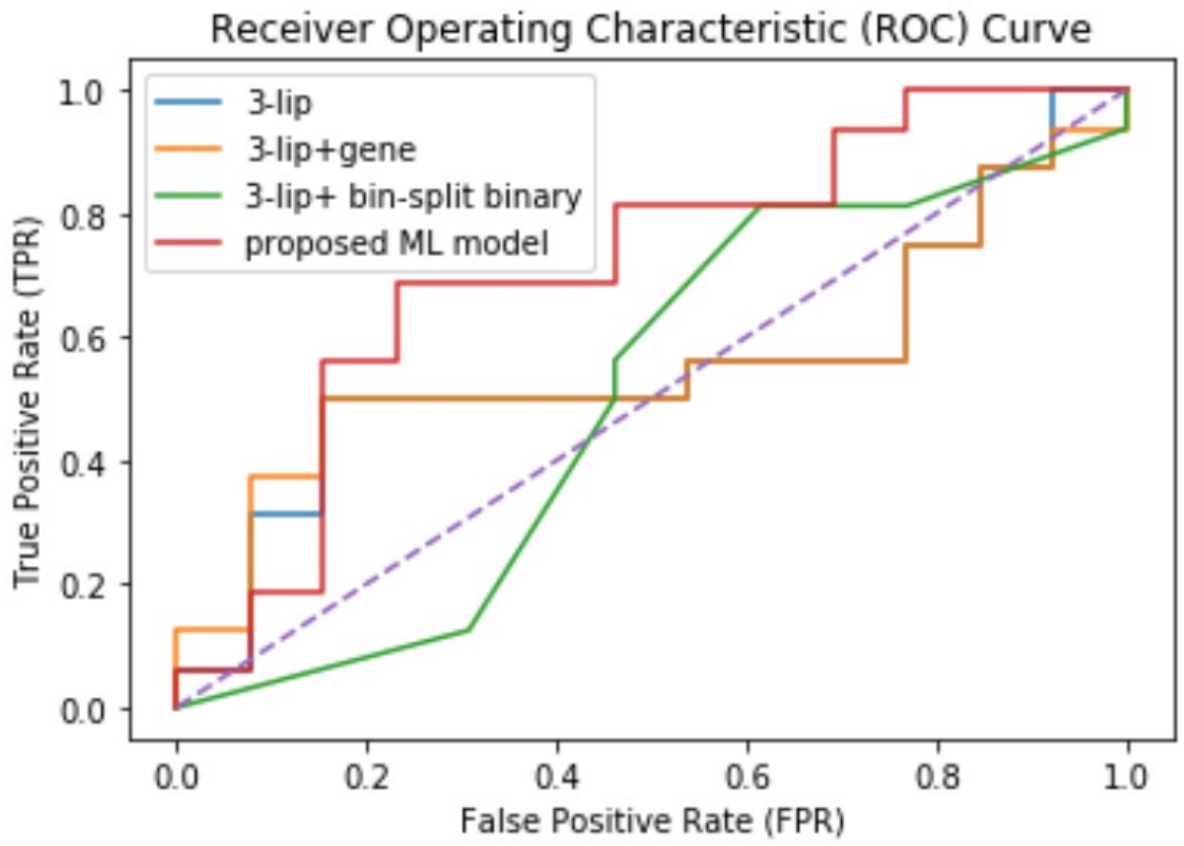


Figure S2

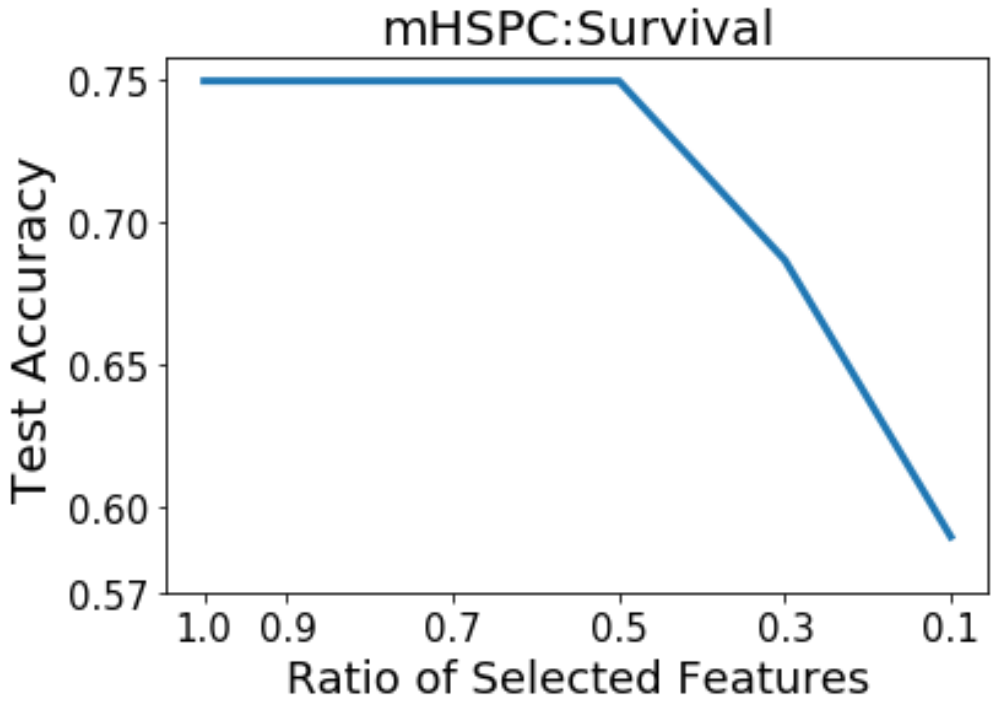


Figure S3

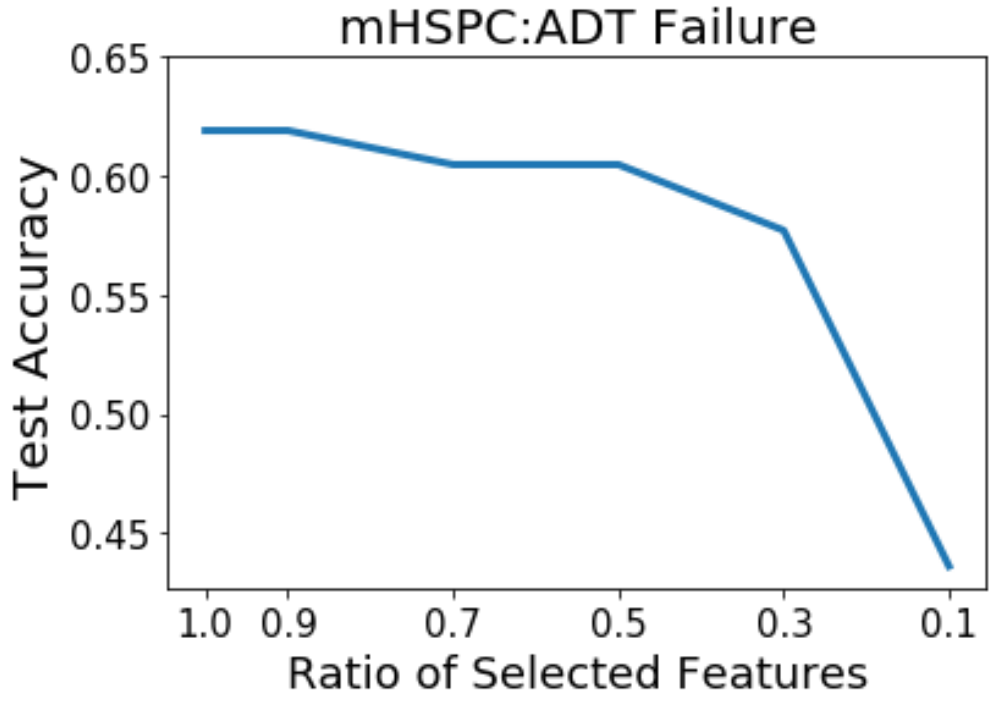


Figure S4

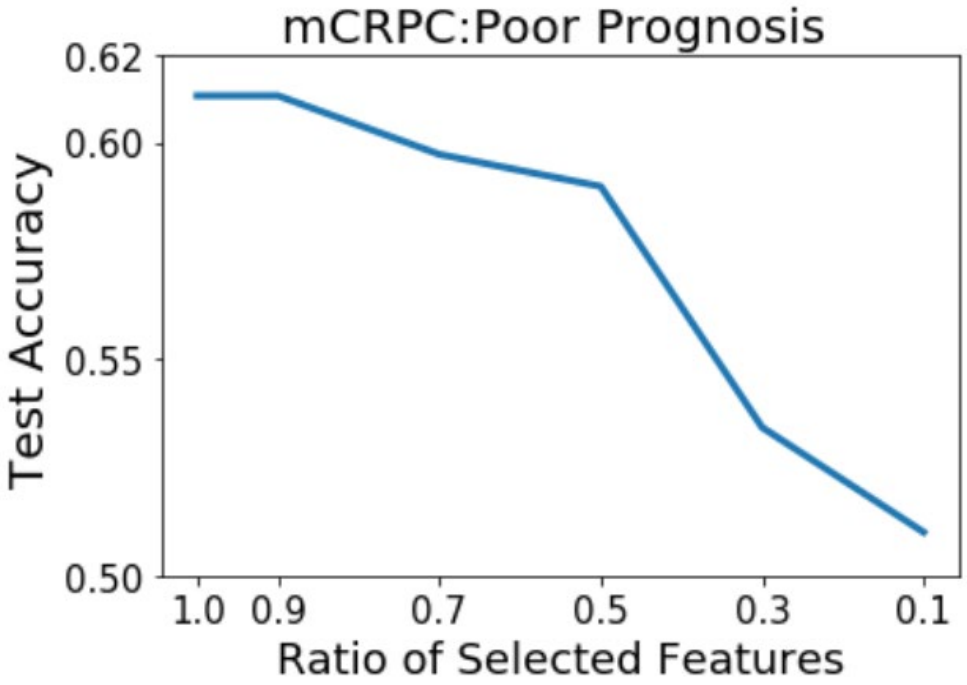


Figure S5

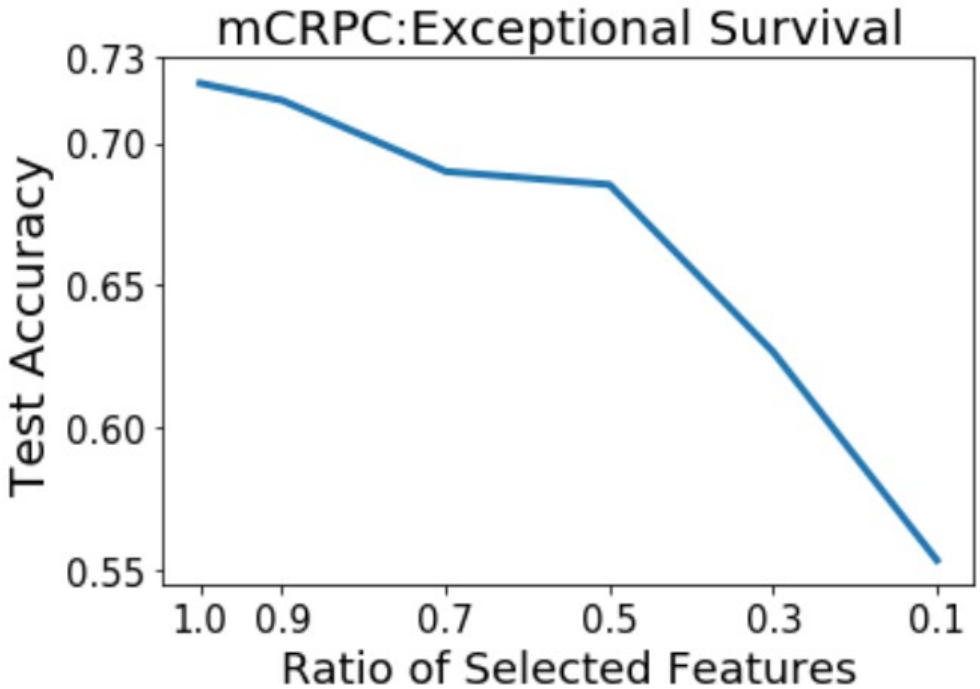


Figure S6

- Given a model trained for mHSPC ADT failure with feature weights:
  (assume there are only 3 features for simplicity)

Feature name	gene_CHECK2_Mut	lip_DG(16:18)_high	lip_DG(16:18)_low
Feature weight	-0.52	0.68	-0.65

- Assume there are 2 new patients with following features values

Feature name/values	gene_CHECK2_Mut	lip_DG(16:18)_high	lip_DG(16:18)_low
Patient #1	0	1	0
Patient #2	1	0	1

- Then we combine their features values and corresponding

  Score(Pt #1 ADT will fail) = 1/(1+exp(-(0*0.52 + 1*0.68 + 0*-0.65))) = 0.66 > 0.5 => predict "ADT will fail"

  Score(Pt #2 ADT will fail) = 1/(1+exp(-(1*0.52 + 0*0.68 + 1*-0.65))) = 0.24 < 0.5 => predict "ADT will not fail"