Optimizing transcriptome-based synthetic lethality predictions to improve precision oncology in early-stage breast cancer: BC-SELECT

Yewon Kim[1],
Matthew Nagy[2],
Becca Pollard[1],
Padma Sheila Rajagopal[1,3]

[1] Cancer and Data Science Laboratory, Center for Cancer Research, National Cancer Institute, Bethesda, MD
[2] Boston Children's Hospital, Boston, MD
[3] Women's Malignancies Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD

Abstract Word Count: 285 words
Word Count: 3,621 words
References: 58 references

Figure / Table Count:
Figures: 4
Tables: 2
Supplementary Figures: 2
Supplementary Tables: 3

ABSTRACT

Introduction: Patients with early-stage breast cancer, especially in the triple-negative and HER2-positive subtypes, do not yet have approaches to predict relative benefit from therapies before treatment. SELECT[1] is a computational tool that leverages genetic interactions, such as synthetic lethality or rescue (SL or SR), to predict treatment response from tumor transcriptomics data. SELECT was tested on endocrine therapy and HER2-directed therapy, leaving open the application to treatment options used in triple-negative breast cancer, such as PARP inhibitors or immunotherapy. We adapt SELECT to address this clinical context, broadening the application of transcriptome-based genetic interactions in patients with early-stage breast cancer.

Methods: SELECT involves two main steps: (1) a “training” step that, for a given drug target, constructs a library of clinically relevant candidate SL/SR partner genes, and (2) a “testing/validation” step that calculates patients' predicted response scores in unseen studies based on the relative expression of the candidate partner genes. In our approach, BC-SELECT, we apply breast cancer-specific expression data from breast cancer trials and datasets to both the training and testing steps.

Results: BC-SELECT predicted responders and non-responders in 11 clinical trials in breast cancer not used for training, with predictions for 7 of 11 trials (including all PARP inhibitor and immunotherapy trials) demonstrating statistically significant odds ratios using multiplicity-corrected Fisher exact tests. BC-SELECT not only outperformed its

score components (when each was compared to chance), but also outperformed SELECT.

Conclusion: In the early-stage setting, particularly given that there are no pre-treatment predictors of response in triple-negative and HER2-positive breast cancer, BC-SELECT predicts patients anticipated to respond to PARP inhibitors and immunotherapy. These findings underscore the algorithm's potential as a decision-support tool to prioritize treatments for patients with early-stage breast cancer.

INTRODUCTION

Gene expression data (via RT-PCR-based panels) have transformed clinical management in early-stage breast cancer. However, this information is currently limited to patients with hormone receptor positive (HR+) disease, who benefit from the opportunity to undergo this testing prior to receiving chemotherapy[2–9]. Patients with HER2-positive (HER2+) and triple-negative breast cancer (TNBC), unfortunately, must currently undergo multiple therapies and surgery, with pathologic complete response (pCR) to treatment serving as a post-facto indicator of response. Although broader transcriptomic data, which can be routinely obtained commercially, identifies not only gene expression levels, but also fusion genes that underlie FDA therapeutic approvals, this data is not used routinely for early breast cancer management[10–12].

Synthetic lethality (SL), a form of genetic interaction in which two genes must be concurrently inactivated to cause cell death, demonstrates clinical relevance, as shown in the example of PARP inhibitors as used to treating cancers with a *BRCA1/2* pathogenic mutation across numerous clinical trials[13,14]. Other genetic interactions, such as synthetic dosage lethality (SDL), in which one gene must be over-expressed and another under-expressed[15], or synthetic rescue (SR), in which inactivation of one gene that would cause death is ameliorated by altered activation of another gene, are being actively studied for potential clinical applications[16].

Computational approaches to identifying these interactions have been recently developed. SELECT[1] and a commercial adaptation, ENLIGHT[17], use large-scale cell line and patient tumor transcriptomic data to identify clinically relevant synthetic lethal/rescue partners for a wide array of targeted therapies. Both algorithms show

efficacy in predicting treatment response in retrospective clinical trials[1,17]. Despite the breadth of these tools across numerous cancer types as published, their translational application in breast cancer was limited. The breast cancer trials that were included focused on endocrine therapy or HER2-directed therapy, leaving open the application to treatment options used in triple-negative breast cancer, such as PARP inhibitors or immunotherapy.

In this paper, we present an adaptation of SELECT for breast cancer with the goal of predicting therapeutic response to targeted therapy and immunotherapy in early-stage breast cancer. We used datasets from patients with breast cancer to train our model, and we apply the algorithm to clinical trials not only of HER2-directed therapy, but also immunotherapy and PARP inhibitors in breast cancer. This adaptation, BC-SELECT, advances a translational, transcriptome-based strategy to predict treatment responders, with the eventual goal of personalizing prioritized treatment strategies for patients with breast cancer.

RESULTS

Overview of the BC-SELECT framework and adaptations to predict response to targeted therapy and immunotherapy in early-stage breast cancer

Figure 1 provides a visual overview of the BC-SELECT pipeline. There are two parts: training and testing/validation. The training part uses large-scale cell-line and patient datasets to identify clinically relevant genetic interactions for gene pairs A and B (in which A is the target(s) for targeted therapy or immunotherapy, and multiple partner Bs are identified). Expression levels of partner B genes from the testing part are used to generate a patient's predicted response score to a treatment targeting gene A in the testing/validation part, with all patients analyzed only within a given trial dataset. We use separate modules for targeted therapies and for immunotherapies (see Methods). We tested BC-SELECT on 11 breast cancer clinical trials: four using trastuzumab and one using lapatinib in HER2+ disease, two using durvalumab in TNBC, one using pembrolizumab in TNBC and HR+ breast cancer, and three using PARP inhibitors (either veliparib or olaparib) in TNBC and HR+ disease. This approach allows BC-SELECT to avoid overfitting issues while achieving flexibility across datasets.

Briefly, we made the following adaptations to the previously published SELECT algorithm (see Methods):

• The training part uses breast cancer-specific patient datasets, and the testing/validation part uses only breast cancer-specific trials. As all trials were in early-stage TNBC, pCR was used as the response indicator.

* We use breast cancer trials during the testing part (not used in the validation part) to optimize tuning parameters.

• We remove copy number data from the training part to facilitate future addition of publicly available breast cancer datasets.

• We add additional clinical covariates to the survival analyses in the Cox proportional hazards filtering step to improve clinical utility.

• We add a separate SDL score to the targeted therapy prediction module, in which a higher SDL score indicates a higher likelihood of cancer cell death and drug response.

• In the targeted therapy prediction module, we prioritize candidate partner genes based on survival analysis.

• We generate a final predicted response score (described below) that incorporates target gene expression of the tested therapy as well as a ratio of the relative expression of the partner genes for the relevant genetic interaction to the relative expression of all genes.

BC-SELECT predicts responders and non-responders to targeted therapies and immunotherapy across 11 clinical trials

Across all datasets, responders (pCR = 1) had consistently higher predicted scores from BC-SELECT compared to non-responders (pCR = 0). We derive odds ratios of accurately distinguishing between responders and non-responders based on the BC-SELECT final predicted response scores (Figure 2). Odds ratios were statistically significant in 7/11 trials, including all PARP inhibitor and all immunotherapy trials. Odds ratios, 95% confidence intervals, and multiplicity-corrected p-values from Fisher exact

tests for statistical significance are shown in Table 1. Furthermore, using an adjusted p-value threshold of 0.1 after correcting for multiple hypothesis testing, we also observed a statistically significant difference between the predicted response scores in responders and in non-responders in 8/11 trials, again including all of the PARP inhibitor (Figure 3a) and immunotherapy trials (Figure 3b), as determined by the Wilcoxon rank-sum test.

The predicted response score used in BC-SELECT also outperforms its components. Table 2 reviews the ROC-AUC values and empirical p-values across the 11 validation clinical trials for the BC-SELECT predicted response score and its two components: target gene expression and partner gene expression ratio. BC-SELECT predicted response scores were statistically significant in predicting response compared to chance (one-tailed empirical p-value < 0.025) in 4 trials (2/3 trials with a PARP inhibitor and 2/4 trastuzumab trials), with AUCs ranging from 0.64-0.81. Partner gene expression ratio statistically significant in 3 datasets, and target gene expression in 2 datasets. With immunotherapy, BC-SELECT and partner gene expression ratio were both statistically significant in predicting response compared to chance in all datasets, but ROC-AUCs were consistently higher with BC-SELECT (0.74-0.76 vs. 0.65-0.72). Target gene expression was only significant in 2/3 datasets. ROC and precision-recall curves for these datasets, given the limitations of ROC-AUC alone, are shown in Supplementary Figure 1.

BC-SELECT performs better than SELECT and ENLIGHT in early-stage breast cancer immunotherapy datasets

Where original SELECT results were available or could be estimated, BC-SELECT consistently achieved higher ROC-AUCs than SELECT (Figure 4a). We were also able to compare four breast cancer datasets to results generated from ENLIGHT (Figure 4b). BC-SELECT also performed better than ENLIGHT in the one immunotherapy dataset. However, BC-SELECT's performance was mixed relative to ENLIGHT for trastuzumab, with superior performance in only 1/3 datasets.

DISCUSSION

Our computational algorithm, BC-SELECT, effectively predicts responders to PARP inhibitors and immunotherapy in early-stage breast cancer, demonstrating translational application feasibility. We adapted SELECT with breast cancer-specific trials and datasets for both training and testing/validation, added additional clinical covariates as part of the filtering of genetic interaction pairs, and sought to demonstrate with statistical rigor how well BC-SELECT predicts treatment response. Given that the eventual goal would be to support and refine current clinical treatment strategies with clinical utility, we intentionally applied multiple measures with stringent thresholds to test performance. These statistical measures included not only odds ratios but ROC-AUCs, precision-recall curves, and assessments of BC-SELECT and score components in Monte-Carlo simulations (to develop an estimate of “chance”). BC-SELECT distinguished responders from non-responders in seven trials, including all performed with PARP inhibitors and immunotherapy.

Potential future applications of BC-SELECT can encompass multiple clinical needs. One application already alluded to is the ability to predict responders to therapy in HER2+ and TNBC breast cancers in advance of treatment. While HER2+ subtyping with immunohistochemistry and fluorescence in situ hybridization has been long established, research remains ongoing into methods that can improve on these subtyping strategies specifically for treatment response prediction[18]. Another potential area of need is the tool’s ability to identify potential responders to therapy that are outside current guidelines. For example, patients with metastatic TNBC are currently only candidates for immunotherapy if they have a combined *PD-L1* positive expression

score above 10, with an active question being how to identify potential responders if any that do not meet this specific threshold[19]. The structure of BC-SELECT facilitates the addition of metastatic breast cancer datasets in future to study this question.

This tool can also prove valuable in the context of patients who have multiple options potentially available, supporting clinicians in prioritizing given options. Patients with pathogenic germline *BRCA1/2* variants and TNBC, for example, are often eligible for both PARP inhibitors and immunotherapy, with guidelines having limited data to support current recommendations for prioritization. Multiple clinical trials have not demonstrated an outcome benefit to combining these therapies for this patient population[20–24]. This approach could potentially optimize decision making in such scenarios.

TNBC-specific subtyping is a separate translational approach that uses gene expression data to classify breast cancers within TNBCs and identify potential treatment targets[25]. This approach is not competitive to ours, but rather, complementary given that our approach adds predicted genetic interactions. Additionally, TNBC-specific subtyping has not yet demonstrated the ability to prioritize current treatments.

During the development of BC-SELECT, we observed notable translationally relevant patterns. We observed that target gene expression, specifically, the product of *PDCD1* and *CD274*, performed well in predicting response to immunotherapy. *CD274* / *PD-L1* is an existing therapeutic biomarker via protein expression, with part of the challenge in its current predictive use being multiple assays and incomplete ascertainment of potential treatment responders[26]. RNA-based expression of these genes is known to correlate with immunotherapy responders[27–29], but these genes have

not demonstrated pure predictive performance in themselves. In this paper, we still observed that BC-SELECT was improved upon target gene expression performance in immunotherapy, reinforcing that consideration of both the synthetic genetic interaction mechanism (where it applies) as well as the mechanism of the drug support prediction.

We also observed that, given our inclusion of both SL and SDL in the targeted therapy module, therapeutic targets would typically show better performance with genetic one interaction type over another across trials. PARP inhibitors, for instance, demonstrated better performance with SL partners, consistent with their known mechanism of interaction. Interestingly, HER2-targeted therapies showed better performance with SDL partners. HER2 has only been studied as a participant in synthetic lethal interactions to a minimal extent to date[30,31]. Our result encourages further investigation into how HER2-targeted agents may be applied to cancers through this genetic interaction strategy.

BC-SELECT does have limitations consistent with the other computational algorithms in this family. It requires specific drug targets. Many chemotherapies, unfortunately, have numerous target effects by mechanism, limiting the ability to predict chemotherapy response using this tool. If a drug has multiple targets, as of now, we consider the average gene expression of these targets, weighting the targets equally. We observed an issue with this in the dataset GSE66399[32] (lapatinib), a drug that is designated as having both *ERBB2* (HER2) and *EGFR* as targets in DrugBank. If we prioritize *ERBB2* (HER2) instead of equally weighting these targets, BC-SELECT achieved an ROC-AUC of 0.67 (as opposed to 0.51), suggesting that an ability to assign biologically informed weights to the targets would enhance BC-SELECT's prediction

performance. We also observed that datasets with fewer than 25 patients showed reduced prediction accuracy, establishing a lower threshold for application. As more clinical datasets in breast cancer become available for training and testing/validation, we can include these, reinforcing against BC-SELECT's vulnerability to "noise."

In summary, we have developed BC-SELECT, a computational tool designed to predict drug response to targeted and immunotherapy in patients with breast cancer using gene expression and synthetic genetic interactions. This tool maximizes roles for publicly available datasets with human patient data and focuses on clear clinically focused outcomes. Given the increased success rates of SL oncology trials even outside of PARP inhibitors[33], it is worth exploring genetic interactions and their applications across therapeutic options further. This tool demonstrates predictive performance across targeted therapies and immunotherapies, enabling the differentiation of responders and non-responders for future translational purposes and eventual clinical decision support.

METHODS

Datasets used for training

To train BC-SELECT, we use large, publicly available breast cancer-specific datasets that included gene expression data (microarray or normalized RNA-seq), clinical measures (age, stage), and overall survival. This includes all patients with breast cancer in TCGA (n=1075)[34] as well as METABRIC (n=1989)[35]. The cell-line data is identical to the original SELECT.

Drug targets

We use DrugBank as our reference for the drug targets for targeted therapy[36]. For immunotherapy, the "drug targets" were the simple products of the expression of *PDCD1* and *CD274.*

BC-SELECT training framework

Our model adapts the original SELECT training framework[1]. Screening for genetic interactions is performed in two independent modules, one that employs SL/SDL interactions for targeted therapies, and one that employs SR interactions for immunotherapy.

There are three general filtering steps in the training framework to identify gene pairs associated with clinical relevance:

(1) Either a large-scale data screen using tumor cell line data[37–42] and pharmacological data[43–45] with a Wilcoxon rank-sum test to identify significant associations (targeted

therapy) or a hypergeometric test (described in ISLE[46] and SELECT[1]) in the training datasets above to identify pairs of genes that were under expressed or inactivated together more frequently than expected by chance in a hypergeometric distribution.
(2) Cox proportional hazards models using the genetic interaction partners' gene expressions from patients with breast cancer (in the training datasets above) to identify genetic interactions associated with patient survival.
(3) Evolutionary analysis of preserved gene expression relationships using the Euclidean distance of phylogenetic similarity scores[47] and feature weight[48], given the benefit of this filter to prioritizing gene pairs as previously described[46].

Gene pairs advance through filters if they meet the pre-specified FDR threshold of 0.2 for the Wilcoxon rank sum test / hypergeometric test in step (1) and survival analysis in (2). This threshold is used to avoid overly conservative exclusion of potential genetic interactions, and is the same uniform threshold used in ISLE[46]. For the phylogenetic analysis in step (3), we identify phylogenetically linked candidate partner genes by analyzing phylogenetic profiles.

Each module includes tailored adaptations. In the targeted therapy module, step (1) is the large-scale data screen using tumor cell line and pharmacological data as described. For step (2), we conduct a total of four separate Cox proportional hazard analyses: one for SDL and one for SL, each performed separately in the TCGA[34] and METABRIC[35] datasets to avoid batch effect confounding. Clinical covariates include age, stage, sex (for TCGA) and race (for TCGA). Gene expression values are included

as covariates in addition to clinical variables to determine if expression of those genes is significantly associated with patient survival. Gene expression is assessed only within datasets (not across datasets, again to limit batch effect confounding) and split into tertiles (where upregulated is the top tertile, and downregulated is the bottom tertile based on relative expression levels across patients. Significance is determined by using the multiplicity-corrected marginal p-values of covariates for gene expression and the p-values from the associated log-likelihood ratio tests.

After the partner genes are identified, the training framework selects the top 50 SL or SDL partner genes, prioritized by their adjusted significance in the Cox proportional hazards testing. These genes are used to calculate the patient's partner gene expression ratios (Supplementary Table 1).

For the immunotherapy module, for step (1), the cell-line and drug screen filter did not identify enough immune-related candidate partner genes. As such, we substitute a hypergeometric test only for this module, as performed in ISLE[46] and SELECT[1], using the breast TCGA and METABRIC datasets. This test identifies pairs of genes that are under expressed or inactivated together more frequently than expected by chance in a hypergeometric distribution (suggesting that this would support cancer cell survival). For step (2), two survival analyses were performed using the same TCGA and METABRIC datasets. We focused on anti-PD1/PDL1 agents, as other classes of immunotherapy are not yet approved in breast cancer and did not have corresponding clinical trial data available. For step (3), we compute the phylogenetic similarity score between each

candidate partner gene and both *PDCD1* and *CD274*. The final phylogenetic score for the candidate gene is determined as the smaller value between the two similarity scores. We identify phylogenetically linked candidate partner genes using a threshold (acceptable maximum Euclidean distance between two linked genes) of 10.5 for targeted therapy[46] and 23.15 for immunotherapy[1].

After the partner genes are identified, the training framework selects the top 10 SR partner genes, prioritized by phylogenetic similarity scores. These genes are used to calculate the patient's partner gene expression ratios (Supplementary Table 1).

BC-SELECT predicted response scores

The partner gene expression ratio for SL in the targeted therapy module for each patient is calculated as the ratio of down-regulated partner genes (as defined by tertiles) over all 50 SL genes identified. The partner gene expression ratio for SDL in the targeted therapy module for each patient is calculated as the ratio of up-regulated partner genes (as defined by tertiles) over all 50 SDL genes identified. The partner gene expression ratio for SR in the immunotherapy module is calculated as 1 minus the ratio of down-regulated partner genes over all 10 SR genes identified (as we are measuring the likelihood of drug response).

These gene expression ratios are then revised to include only those genes in both the numerator and denominator that are in curated gene lists of 776 genes (breast cancer-specific) and 1154 genes (immune-specific) derived from Nanostring[49,50]

(Supplementary Table 2) for normalization and noise reduction. The final predicted response scores were calculated by multiplying the partner gene expression score by target gene expression. Where a drug has multiple targets, the score uses the mean expression value of all targets for each patient. These predicted response scores are what are reported as the predicted response to targeted therapy or immunotherapy.

BC-SELECT testing and validation framework

Clinical trials used for testing and validation

To test and validate BC-SELECT, we sought publicly available clinical trial datasets that included at least 25 patients with early-stage breast cancer, individual level gene expression data, and treatment response. pCR was the most studied trial endpoint available. This was consistently reported across studies as a binary measure with either 0 (residual disease) or 1 (complete response).

Three datasets (two for targeted therapy and one for immunotherapy) are used as tuning cohorts to determine the hyperparameters. These fixed parameters are then used for drug response predictions in independent targeted and immunotherapy datasets to avoid overfitting. We included a total of 8 clinical trials across 4 therapies for targeted therapy and 3 clinical trials for 2 therapies for immunotherapy. Trials used in this paper are described in Supplementary Table 3.

Performance evaluation

To test the performance of BC-SELECT, we compared the final prediction scores of BC-SELECT in response to a given therapy to the actual drug responses within each clinical trial. We calculated odds ratios, ROC-AUC and precision-recall curves to assess prediction performance.

For the odds ratio of accurate response prediction, we chose a classification threshold of 0.5 due to the binary nature of the response metric. Patients with a final predicted BC-SELECT score greater than 0.5 were classified as predicted responders, while those with a score of 0.5 or lower were classified as predicted non-responders. Using the true drug response and predicted response status, we calculated the odds ratios along with the corresponding 95% confidence intervals.

We estimated the statistical significance of BC-SELECT's predictions compared to chance by developing a prediction "null distribution" using Monte-Carlo simulations (Supplementary Figure 2). For the Monte-Carlo simulations, we randomly selected a subset of genes with the same number of partners as our candidates (50 for targeted therapy and 10 for immunotherapy). Using these genes, we calculated ROC-AUC values and iterated this process 5000 times to yield a distribution of "null AUC values". We then could compare the predictive efficacy of the BC-SELECT final prediction score via percentile of the ROC-AUC to that achieved by chance. We used a similar approach to generate a separate "null" distribution in order to compare the predictive performance of the gene expression ratios to chance. For target gene expression, we used the same

number of randomly selected genes as number of targets for the drug of interest to generate a separate “null” distribution. We also report our findings using an empirical p-value, the probability of the null AUC distribution exceeding the obtained ROC-AUC of BC-SELECT. Based on empirical rules, the top 2.5% (statistical significance threshold p=0.025), corresponds to the single tail with prediction better than two standard deviations from the mean in the null distribution.

The final reported odds ratios and ROC-AUC values for targeted therapy represented either all SL or all SDL gene pairs – they were independently generated and compared within the targeted therapy module.

Statistical Analysis

Statistical tests used to prioritize partner genes were described in “**BC-SELECT training framework.”** Cox proportional hazards regression models were fitted using the R library survival. Wilcoxon rank sum tests were performed using R functions wilcox.test. To control the false discovery rate (FDR), we apply the Benjamini–Hochberg procedure (BH step-up procedure) as a multiplicity correction using p.adjust function in R. For the performance evaluation, The Fisher exact test was performed using the fisher.test function in R. ROC-AUC values are calculated using the R library ROCR and pROC. Precision-recall curves were generated using the R libraries pROC and ggplot2. All analyses were conducted using version 4.2 of the R software.

References

1. Lee JS, Nair NU, Dinstag G, et al. Synthetic lethality-mediated precision oncology via the tumor transcriptome. *Cell*. 2021;184(9):2487-2502.e13. doi:10.1016/j.cell.2021.03.030

2. Prat A, Ellis MJ, Perou CM. Practical implications of gene-expression-based assays for breast oncologists. *Nat Rev Clin Oncol*. 2011;9(1):48-57. doi:10.1038/nrclinonc.2011.178

3. Kalinsky K, Barlow WE, Gralow JR, et al. 21-Gene Assay to Inform Chemotherapy Benefit in Node-Positive Breast Cancer. *N Engl J Med*. 2021;385(25):2336-2347. doi:10.1056/NEJMoa2108873

4. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med*. 2016;375(8):717-729. doi:10.1056/NEJMoa1602253

5. Sparano JA, Gray RJ, Makower DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med*. 2018;379(2):111-121. doi:10.1056/NEJMoa1804710

6. Piccart M, van 't Veer LJ, Poncet C, et al. 70-gene signature as an aid for treatment decisions in early breast cancer: updated results of the phase 3 randomised MINDACT trial with an exploratory analysis by age. *Lancet Oncol*. 2021;22(4):476-488. doi:10.1016/S1470-2045(21)00007-3

7. Nielsen T, Wallden B, Schaper C, et al. Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens. *BMC Cancer*. 2014;14:177. doi:10.1186/1471-2407-14-177

8. Dubsky P, Brase JC, Jakesz R, et al. The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2- breast cancer patients. *Br J Cancer*. 2013;109(12):2959-2964. doi:10.1038/bjc.2013.671

9. O'Regan RM, Zhang Y, Fleming GF, et al. Breast Cancer Index in Premenopausal Women With Early-Stage Hormone Receptor-Positive Breast Cancer. *JAMA Oncol*. Published online August 15, 2024:e243044. doi:10.1001/jamaoncol.2024.3044

10. Mundi PS, Dela Cruz FS, Grunn A, et al. A Transcriptome-Based Precision Oncology Platform for Patient-Therapy Alignment in a Diverse Set of Treatment-Resistant Malignancies. *Cancer Discov*. 2023;13(6):1386-1407. doi:10.1158/2159-8290.CD-22-1020

11. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) Breast Cancer v.4.2024. https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf

12. Priedigkeit N, Lebrón-Torres A, Liao J, et al. Abstract GS03-09: Characterization and proposed therapeutic exploitation of fusion RNAs in metastatic breast cancers. *Cancer Research*. 2024;84(9_Supplement):GS03-09. doi:10.1158/1538-7445.SABCS23-GS03-09

13. Ashworth A, Lord CJ. Synthetic lethal therapies for cancer: what's next after PARP inhibitors? *Nat Rev Clin Oncol*. 2018;15(9):564-576. doi:10.1038/s41571-018-0055-6

14. O'Neil NJ, Bailey ML, Hieter P. Synthetic lethality and cancer. *Nat Rev Genet*. 2017;18(10):613-623. doi:10.1038/nrg.2017.47

15. Megchelenbrink W, Katzir R, Lu X, Ruppin E, Notebaart RA. Synthetic dosage lethality in the human metabolic network is highly predictive of tumor growth and cancer patient survival. *Proc Natl Acad Sci U S A*. 2015;112(39):12217-12222. doi:10.1073/pnas.1508573112

16. Sahu AD, S Lee J, Wang Z, et al. Genome-wide prediction of synthetic rescue mediators of resistance to targeted and immunotherapy. *Mol Syst Biol*. 2019;15(3):e8323. doi:10.15252/msb.20188323

17. Dinstag G, Shulman ED, Elis E, et al. Clinically oriented prediction of patient response to targeted and immunotherapies from the tumor transcriptome. *Med*. 2023;4(1):15-30.e8. doi:10.1016/j.medj.2022.11.001

18. Katayama A, Miligy IM, Shiino S, et al. Predictors of pathological complete response to neoadjuvant treatment and changes to post-neoadjuvant HER2 status in HER2-positive invasive breast cancer. *Mod Pathol*. 2021;34(7):1271-1281. doi:10.1038/s41379-021-00738-5

19. Cortes J, Rugo HS, Cescon DW, et al. Pembrolizumab plus Chemotherapy in Advanced Triple-Negative Breast Cancer. *N Engl J Med*. 2022;387(3):217-226. doi:10.1056/NEJMoa2202809

20. Cussac AL, Rugo HS, Robson ME, et al. 198P Pembrolizumab plus olaparib vs pembrolizumab plus chemotherapy after induction with pembrolizumab plus chemotherapy for locally recurrent inoperable or metastatic TNBC: Patient-reported outcomes from KEYLYNK-009. *ESMO Open*. 2024;9:103220. doi:10.1016/j.esmoop.2024.103220

21. Thawani R, Kummar S. Combining PARP Inhibitor With Immunotherapy-Does the Promise of Preclinical Data Translate to Clinic? *JAMA Oncol*. 2023;9(1):25-27. doi:10.1001/jamaoncol.2022.4591

22. Domchek SM, Postel-Vinay S, Im SA, et al. Olaparib and durvalumab in patients with germline BRCA-mutated metastatic breast cancer (MEDIOLA): an open-label, multicentre, phase 1/2, basket study. *Lancet Oncol*. 2020;21(9):1155-1164. doi:10.1016/S1470-2045(20)30324-7

23. Cortesi L, Venturelli M, Cortesi G, et al. A phase II study of pembrolizumab plus carboplatin in BRCA-related metastatic breast cancer (PEMBRACA). *ESMO Open*. 2023;8(2):101207. doi:10.1016/j.esmoop.2023.101207

24. Maintenance Pembrolizumab/Olaparib Misses Survival End Points in Locally Advanced or Metastatic TNBC. OncLive. December 6, 2023. Accessed September 1, 2024. https://www.onclive.com/view/maintenance-pembrolizumab-olaparib-misses-survival-end-points-in-locally-advanced-or-metastatic-tnbc

25. Lehmann BD, Pietenpol JA, Tan AR. Triple-negative breast cancer: molecular subtypes and new targets for therapy. *Am Soc Clin Oncol Educ Book*. Published online 2015:e31-39. doi:10.14694/EdBook_AM.2015.35.e31

26. Grossman JE, Vasudevan D, Joyce CE, Hildago M. Is PD-L1 a consistent biomarker for anti-PD-1 therapy? The model of balstilimab in a virally-driven tumor. *Oncogene*. 2021;40(8):1393-1395. doi:10.1038/s41388-020-01611-6

27. Conroy JM, Pabla S, Nesline MK, et al. Next generation sequencing of PD-L1 for predicting response to immune checkpoint inhibitors. *J Immunother Cancer*. 2019;7(1):18. doi:10.1186/s40425-018-0489-5

28. Kang SY, Heo YJ, Kwon GY, Kim KM. Expression of CD274 mRNA Measured by qRT-PCR Correlates With PD-L1 Immunohistochemistry in Gastric and Urothelial Carcinoma. *Front Oncol*. 2022;12:856444. doi:10.3389/fonc.2022.856444

29. Paré L, Pascual T, Seguí E, et al. Association between PD1 mRNA and response to anti-PD1 monotherapy across multiple cancer types. *Ann Oncol*. 2018;29(10):2121-2128. doi:10.1093/annonc/mdy335

30. Ding Y, Gong C, Huang D, et al. Synthetic lethality between HER2 and transaldolase in intrinsically resistant HER2-positive breast cancers. *Nat Commun*. 2018;9(1):4274. doi:10.1038/s41467-018-06651-x

31. Ku AA, Hu HM, Zhao X, et al. Integration of multiple biological contexts reveals principles of synthetic lethality that affect reproducibility. *Nat Commun*. 2020;11(1):2375. doi:10.1038/s41467-020-16078-y

32. Guarneri V, Dieci MV, Frassoldati A, et al. Prospective Biomarker Analysis of the Randomized CHER-LOB Study Evaluating the Dual Anti-HER2 Treatment With Trastuzumab and Lapatinib Plus Chemotherapy as Neoadjuvant Therapy for HER2-Positive Breast Cancer. *Oncologist*. 2015;20(9):1001-1010. doi:10.1634/theoncologist.2015-0138

33. Schäffer AA, Chung Y, Kammula AV, Ruppin E, Lee JS. A systematic analysis of the landscape of synthetic lethality-driven precision oncology. *Med*. 2024;5(1):73-89.e9. doi:10.1016/j.medj.2023.12.009

34. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013;45(10):1113-1120. doi:10.1038/ng.2764

35. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346-352. doi:10.1038/nature10983

36. Knox C, Wilson M, Klinger CM, et al. DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Res*. 2024;52(D1):D1265-D1275. doi:10.1093/nar/gkad976

37. Aguirre AJ, Meyers RM, Weir BA, et al. Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov*. 2016;6(8):914-929. doi:10.1158/2159-8290.CD-16-0154

38. Cheung HW, Cowley GS, Weir BA, et al. Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proc Natl Acad Sci U S A*. 2011;108(30):12372-12377. doi:10.1073/pnas.1109363108

39. Cowley GS, Weir BA, Vazquez F, et al. Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. *Sci Data*. 2014;1:140035. doi:10.1038/sdata.2014.35

40. Marcotte R, Brown KR, Suarez F, et al. Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer Discov*. 2012;2(2):172-189. doi:10.1158/2159-8290.CD-11-0224

41. Marcotte R, Sayad A, Brown KR, et al. Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. *Cell*. 2016;164(1-2):293-309. doi:10.1016/j.cell.2015.11.062

42. Tsherniak A, Vazquez F, Montgomery PG, et al. Defining a Cancer Dependency Map. *Cell*. 2017;170(3):564-576.e16. doi:10.1016/j.cell.2017.06.010

43. Barretina J, Caponigro G, Stransky N, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483(7391):603-607. doi:10.1038/nature11003

44. Basu A, Bodycombe NE, Cheah JH, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell*. 2013;154(5):1151-1161. doi:10.1016/j.cell.2013.08.003

45. Iorio F, Knijnenburg TA, Vis DJ, et al. A Landscape of Pharmacogenomic Interactions in Cancer. *Cell*. 2016;166(3):740-754. doi:10.1016/j.cell.2016.06.017

46. Lee JS, Das A, Jerby-Arnon L, et al. Harnessing synthetic lethality to predict the response to cancer treatment. *Nat Commun*. 2018;9(1):2546. doi:10.1038/s41467-018-04647-1

47. Tabach Y, Golan T, Hernández-Hernández A, et al. Human disease locus discovery and mapping to molecular pathways through phylogenetic profiling. *Mol Syst Biol*. 2013;9:692. doi:10.1038/msb.2013.50

48. Martin FJ, Amode MR, Aneja A, et al. Ensembl 2023. *Nucleic Acids Res*. 2023;51(D1):D933-D941. doi:10.1093/nar/gkac958

49. nCounter Breast Cancer 360. NanoString. 2024. Accessed July 8, 2024. https://nanostring.com/products/ncounter-assays-panels/oncology/breast-cancer-360/

50. PanCancer Immune Profiling. NanoString. 2024. Accessed July 8, 2024. https://nanostring.com/products/ncounter-assays-panels/oncology/pancancer-immune-profiling/

51. Prat A, Bianchini G, Thomas M, et al. Research-based PAM50 subtype predictor identifies higher responses and improved survival outcomes in HER2-positive breast cancer in the NOAH study. *Clin Cancer Res*. 2014;20(2):511-521. doi:10.1158/1078-0432.CCR-13-0239

52. Liu JC, Voisin V, Bader GD, et al. Seventeen-gene signature from enriched Her2/Neu mammary tumor-initiating cells predicts clinical outcome for human HER2+:ERα- breast cancer. *Proc Natl Acad Sci U S A*. 2012;109(15):5832-5837. doi:10.1073/pnas.1201105109

53. Shen K, Qi Y, Song N, et al. Cell line derived multi-gene predictor of pathologic response to neoadjuvant chemotherapy in breast cancer: a validation study on US Oncology 02-103 clinical trial. *BMC Med Genomics*. 2012;5:51. doi:10.1186/1755-8794-5-51

54. Sammut SJ, Crispin-Ortuzar M, Chin SF, et al. Multi-omic machine learning predictor of breast cancer therapy response. *Nature*. 2022;601(7894):623-629. doi:10.1038/s41586-021-04278-5

55. Wolf DM, Yau C, Wulfkuhle J, et al. Redefining breast cancer subtypes to guide treatment prioritization and maximize response: Predictive biomarkers across 10 cancer therapies. *Cancer Cell*. 2022;40(6):609-623.e6. doi:10.1016/j.ccell.2022.05.005

56. Loibl S, O'Shaughnessy J, Untch M, et al. Addition of the PARP inhibitor veliparib plus carboplatin or carboplatin alone to standard neoadjuvant chemotherapy in triple-negative breast cancer (BrighTNess): a randomised, phase 3 trial. *Lancet Oncol*. 2018;19(4):497-509. doi:10.1016/S1470-2045(18)30111-6

57. Pusztai L, Yau C, Wolf DM, et al. Durvalumab with olaparib and paclitaxel for high-risk HER2-negative stage II/III breast cancer: Results from the adaptively randomized I-SPY2 trial. *Cancer Cell*. 2021;39(7):989-998.e5. doi:10.1016/j.ccell.2021.05.009

58. Foldi J, Silber A, Reisenbichler E, et al. Neoadjuvant durvalumab plus weekly nab-paclitaxel and dose-dense doxorubicin/cyclophosphamide in triple-negative breast cancer. *NPJ Breast Cancer*. 2021;7(1):9. doi:10.1038/s41523-021-00219-7

Conflicts of Interest:

Yewon Kim: None
Matthew Nagy: None
Becca Pollard: None
Padma Sheila Rajagopal: None

Acknowledgments:

We wish to thank Dr. Eytan Ruppin, Dr. Lipika Ray, and Dr. Nishanth Nair for their support and counsel throughout the drafting of this manuscript. This work utilized the computational resources of the NIH HPC Biowulf cluster. (http://hpc.nih.gov)

Data availability:

All available data and the source codes are accessible on GitHub: https://github.com/yewon-kim76/BC-SELECT, and the list of references can be found in Zotero: https://www.zotero.org/yewon76/collections/G396YYNB.

Author contributions:

Yewon Kim: Methodology, Software, Validation, Analysis, Writing - Original And Revision. **Becca Pollard**: Data Curation. **Matthew Nagy**: Software, Validation, Writing – Revision. **Padma Sheila Rajagopal**: Conceptualization, Methodology, Validation, Writing - Revision, Resources, Data Curation, Supervision, Administration, Funding.

Tables and Figures

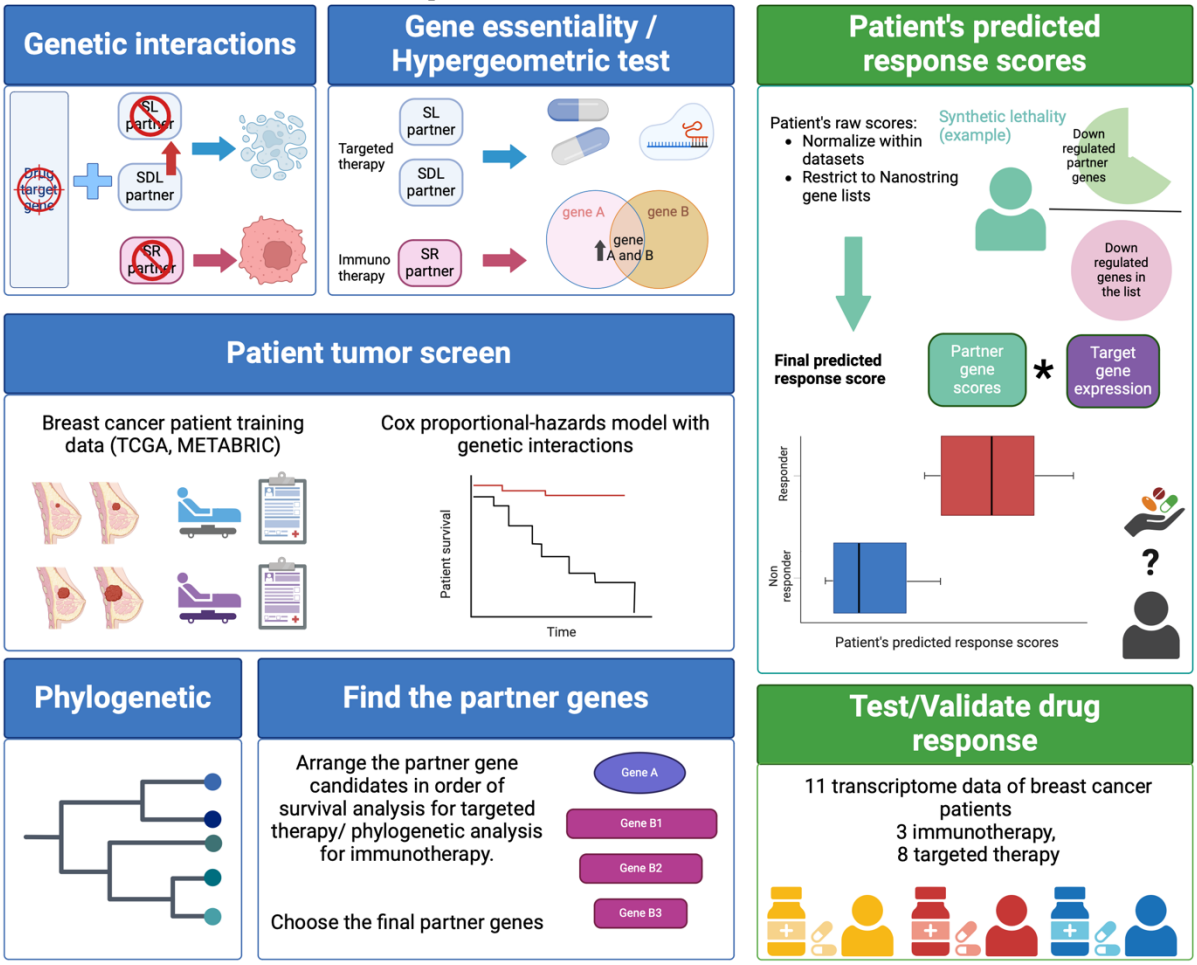


Figure 1: A graphical overview illustrating the adaptation of BC-SELECT pipeline. The training part is depicted in the blue box, while the testing/validation part is illustrated in the green box.

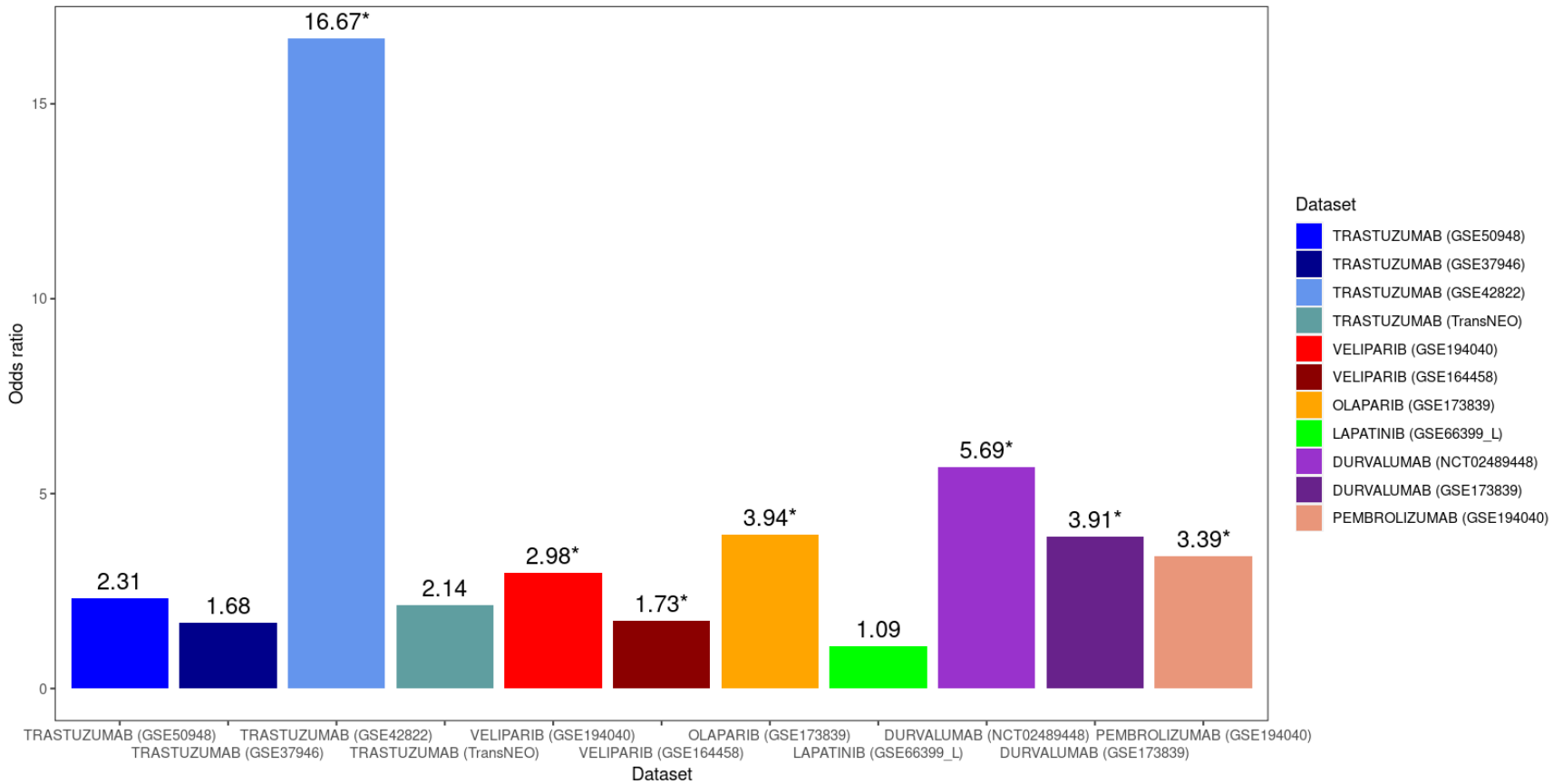
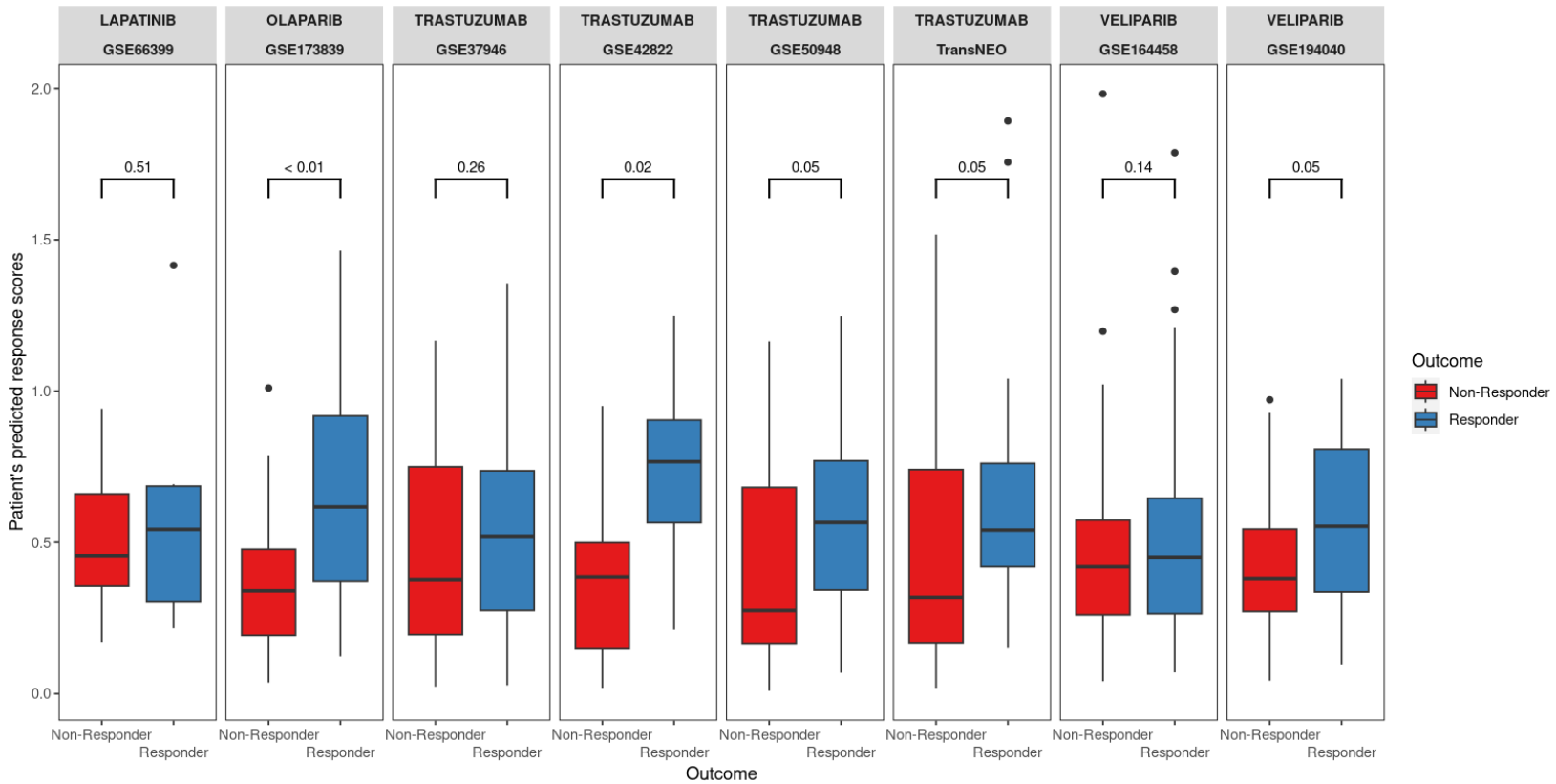


Figure 2: Odds ratios between responders and non-responders using final predicted response scores from BC-SELECT. Using a threshold of 0.5, we calculated the odds ratios between responders and non-responders across 11 breast cancer datasets. All odds ratios were greater than 1, and 7 datasets showed statistical significance (marked with *), with the 95% confidence interval not containing 1. Detailed 95% confidence intervals of odds ratios and multiplicity-corrected p-values from Fisher exact tests are provided in Table 2.

Figure 3

a.


LAPATINIB
GSE66399
OLAPARIB
GSE173839
TRASTUZUMAB
GSE37946
TRASTUZUMAB
GSE42822
TRASTUZUMAB
GSE50948
TRASTUZUMAB
TransNEO
VELIPARIB
GSE164458
VELIPARIB
GSE194040
0.51
< 0.01
0.26
0.02
0.05
0.05
0.14
0.05
Patient's predicted response scores
2.0
1.5
1.0
0.5
0.0
Non-Responder
Responder
Outcome
Outcome
Non-Responder
Responder

b.

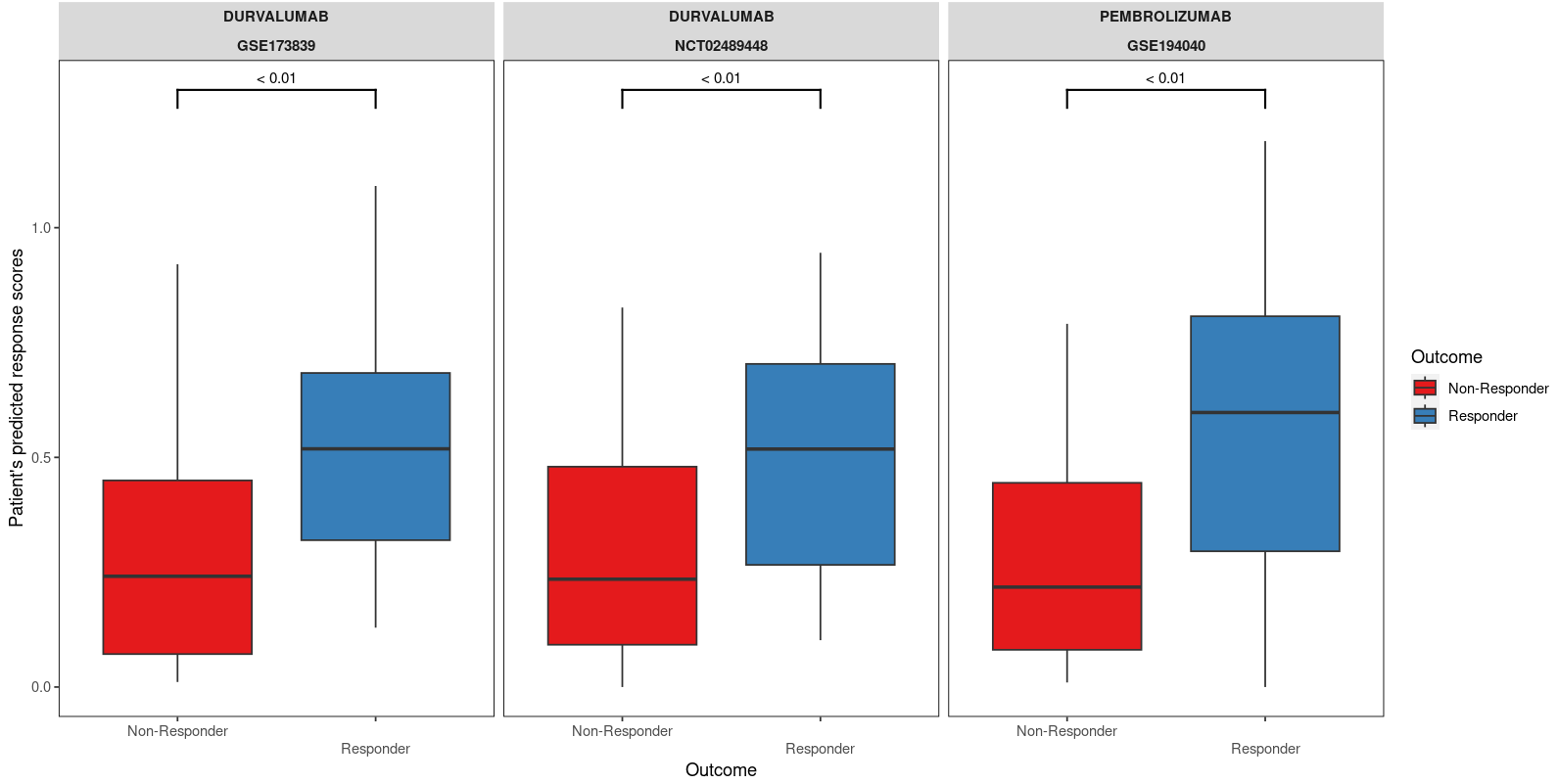
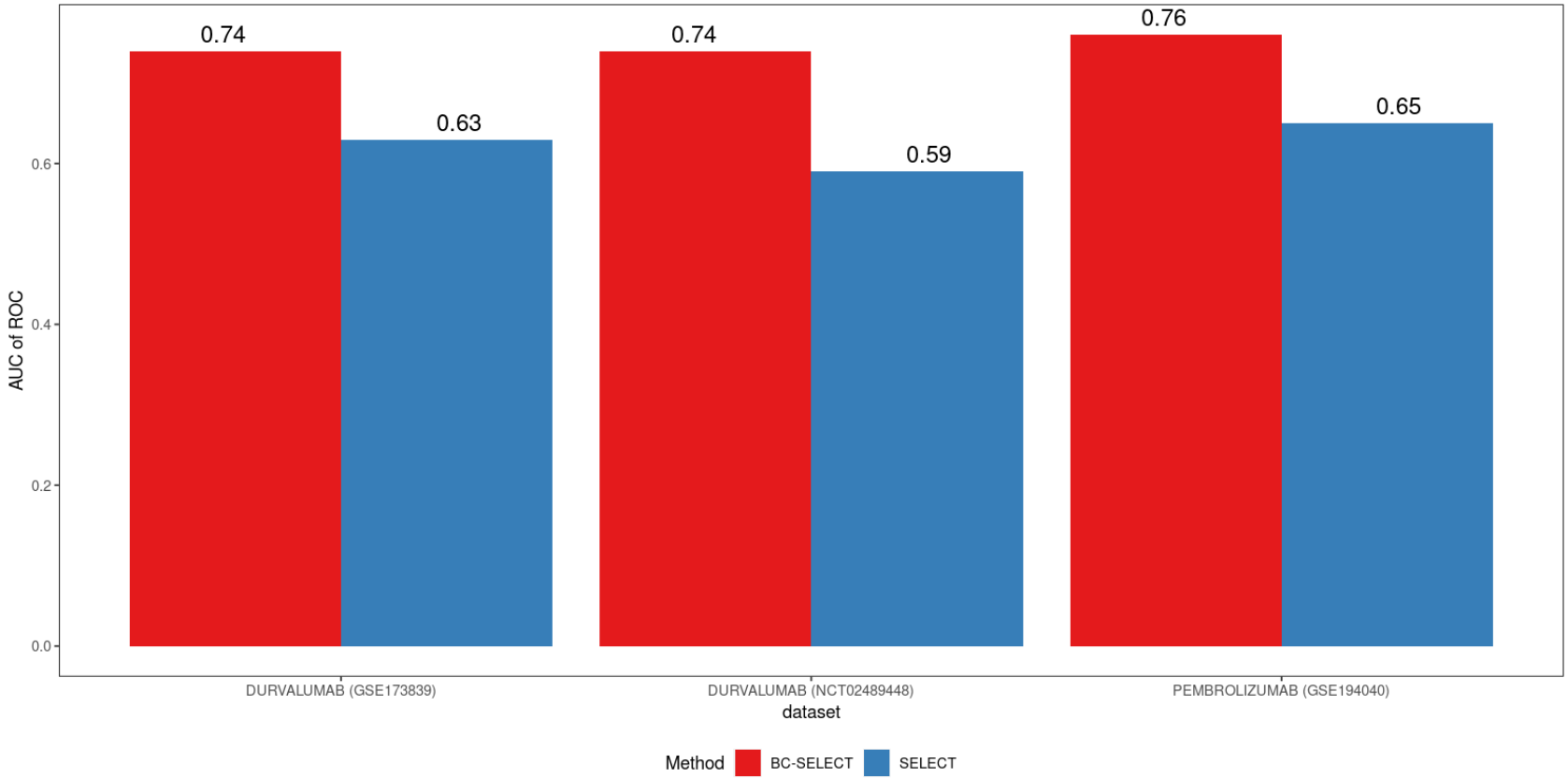


Figure 3: Patient's predicted response scores of BC-SELECT among responders and non-responders. a. Targeted therapies. Final predicted BC-SELECT scores are significantly higher in responders compared to non-responders in 5 out of 8 datasets. **b. Immunotherapy.** Final predicted BC-SELECT scores are significantly higher in responders compared to non-responders in all 3 immunotherapy datasets. Significance was determined via the Wilcoxon rank-sum test with correction for multiple hypothesis testing. Adjusted p-values for the test results are reported in each box plot. Red = non-responders, Blue = Responders.

Figure 4

a.


0.74
0.63
0.74
0.59
0.76
0.65
0.6
0.4
0.2
0.0
AUC of ROC
DURVALUMAB (GSE173839)
DURVALUMAB (NCT02489448)
PEMBROLIZUMAB (GSE194040)
dataset
Method
BC-SELECT
SELECT

b.

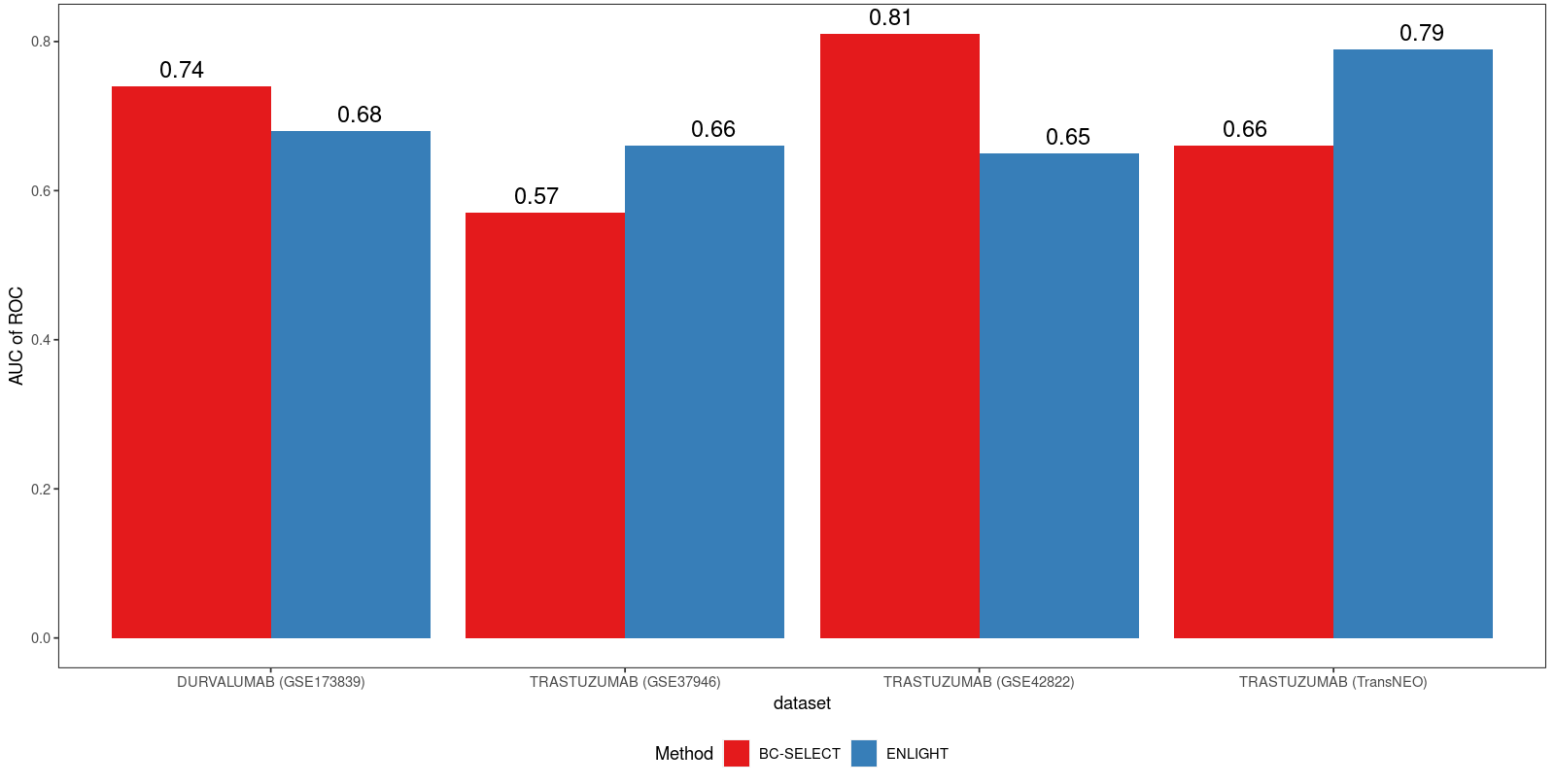


Figure 4: Performance comparisons between BC-SELECT and other algorithms. a. Comparison to SELECT (3 breast cancer immunotherapy datasets): BC-SELECT demonstrates superior prediction performance compared to the SELECT in terms of ROC-AUC values in one pembrolizumab trial and two durvalumab trials.
b. Comparison to ENLIGHT (4 breast cancer datasets): BC-SELECT demonstrates superior prediction performance in the one immunotherapy study, but mixed performance across the trastuzumab studies. Results are reported in ROC-AUC values. Red reflects BC-SELECT, and blue reflects the comparison algorithm (SELECT or ENLIGHT).

	DRUG	TARGET	DATASET	GENETIC INTERACTION	Odds ratio	95% Confidence Interval	Fisher exact test adjusted P-value
Targeted therapy	TRASTUZUMAB	ERBB2 (HER2)	GSE50948[51] (R:31, NR:32)	SDL	2.31	(0.84, 6.34)	0.21
			GSE37946[52] (R:27, NR:23)	SDL	1.68	(0.54, 5.17)	0.47
			GSE42822[53] (R:12, NR:13)	SDL	**16.67**	**(2.27, 122.22)**	**0.04**
			TransNEO[54] (R:19, NR:46)	SL	2.14	(0.72, 6.34)	0.24
	VELIPARIB	PARP1, PARP2	GSE194040[55] (R: 27, NR: 44)	SDL	**2.98**	**(1.10, 8.09)**	**0.09**
			GSE164458[56] (R:127, NR:110)	SL	**1.73**	**(1.02, 2.93)**	**0.09**
	OLAPARIB	PARP1, PARP2, PARP3	GSE173839[57] (R:29, NR:42)	SL	**3.94**	**(1.42, 10.92)**	**0.05**
	LAPATINIB	EGFR, ERBB2 (HER2)	GSE66399[32] (R:8, NR:23)	SDL	1.09	(0.22, 5.45)	1
Immuno therapy	DURVALUMAB	PDCD1, CD274	NCT02489448[58] (R:25, NR:25)	SR	**5.69**	**(1.51, 21.42)**	**0.02**
			GSE173839[57] (R:29, NR:42)	SR	**3.91**	**(1.40, 10.95)**	**0.02**
	PEMBROLIZUMAB		GSE194040[55] (R:31, NR:38)	SR	**3.93**	**(1.39, 11.07)**	**0.02**

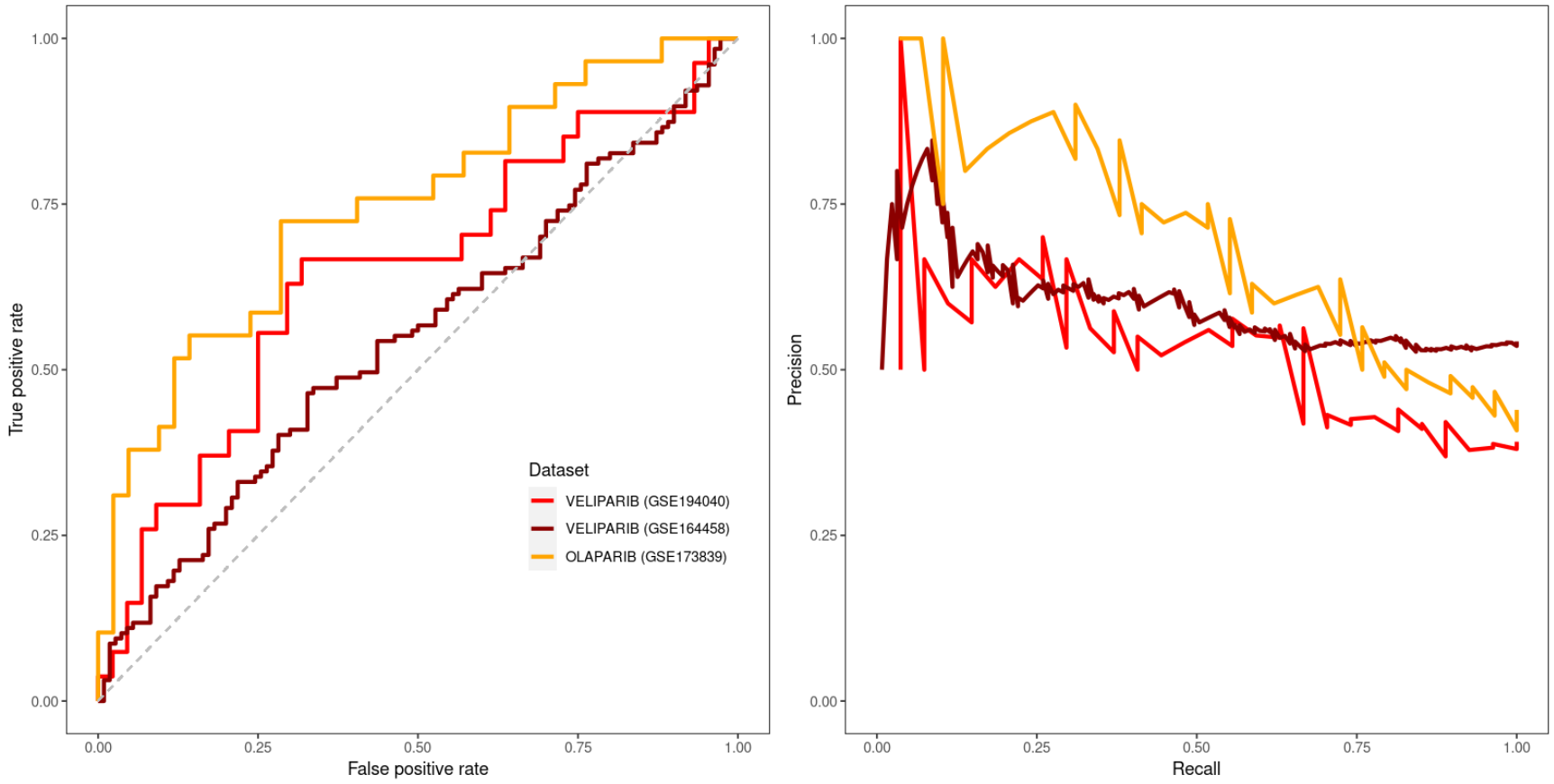
Table 1: Odds ratios between responders and non-responders using patients' predicted response scores from BC-SELECT. All odds ratios were greater than 1, and 7 datasets showed statistical significance (bold). Last column presents the p-values of the Fisher exact test after multiple hypothesis adjustment. R: Responders. NR: Non-Responders. SL: Synthetic Lethality. SDL: Synthetic Dosage Lethality. SR: Synthetic Rescue.

	DRUG	TARGET	DATASET	GENETIC INTERACTION	BC-SELECT ROC-AUC (Empirical P-value)	Partner gene expression ratio ROC-AUC (Empirical P-value)	Target gene expression ROC-AUC (Empirical P-value)
Targeted therapy	TRASTUZUMAB	ERBB2 (HER2)	GSE50948[51] (R:31, NR:32)	SDL	**0.64** (0.01)	**0.63** (0.02)	0.6 (0.06)
			GSE37946[52] (R:27, NR:23)	SDL	0.57 (0.13)	0.44 (0.48)	0.6 (0.1)
			GSE42822[53] (R:12, NR:13)	SDL	**0.81** (0.01)	0.68 (0.23)	**0.78** (0.01)
			TransNEO[54] (R:19, NR:46)	SL	0.66 (0.14)	0.54 (0.64)	0.69 (0.04)
	VELIPARIB	PARP1, PARP2	GSE194040[55] (R: 27, NR: 44)	SDL	**0.65** (0.02)	**0.59** (0.02)	0.63 (0.1)
			GSE164458[56] (R:127, NR:110)	SL	0.55 (0.39)	0.44 (0.99)	**0.61** (0.02)
	OLAPARIB	PARP1, PARP2, PARP3	GSE173839[57] (R:29, NR:42)	SL	**0.75** (<0.01)	**0.79** (<0.01)	0.56 (0.26)
	LAPATINIB	EGFR, ERBB2 (HER2)	GSE66399[32] (R:8, NR:23)	SDL	0.51 (0.53)	0.64 (0.11)	0.43 (0.74)
Immuno therapy	DURVALUMAB	PDCD1, CD274	NCT02489448[58] (R:25, NR:25)	SR	**0.74** (<0.01)	**0.65** (<0.01)	**0.71** (0.01)
			GSE173839[57] (R:29, NR:42)	SR	**0.74** (<0.01)	**0.72** (<0.01)	0.70 (0.03)
	PEMBROLIZUMAB		GSE194040[55] (R:31, NR:38)	SR	**0.76** (<0.01)	**0.67** (<0.01)	**0.75** (<0.01)

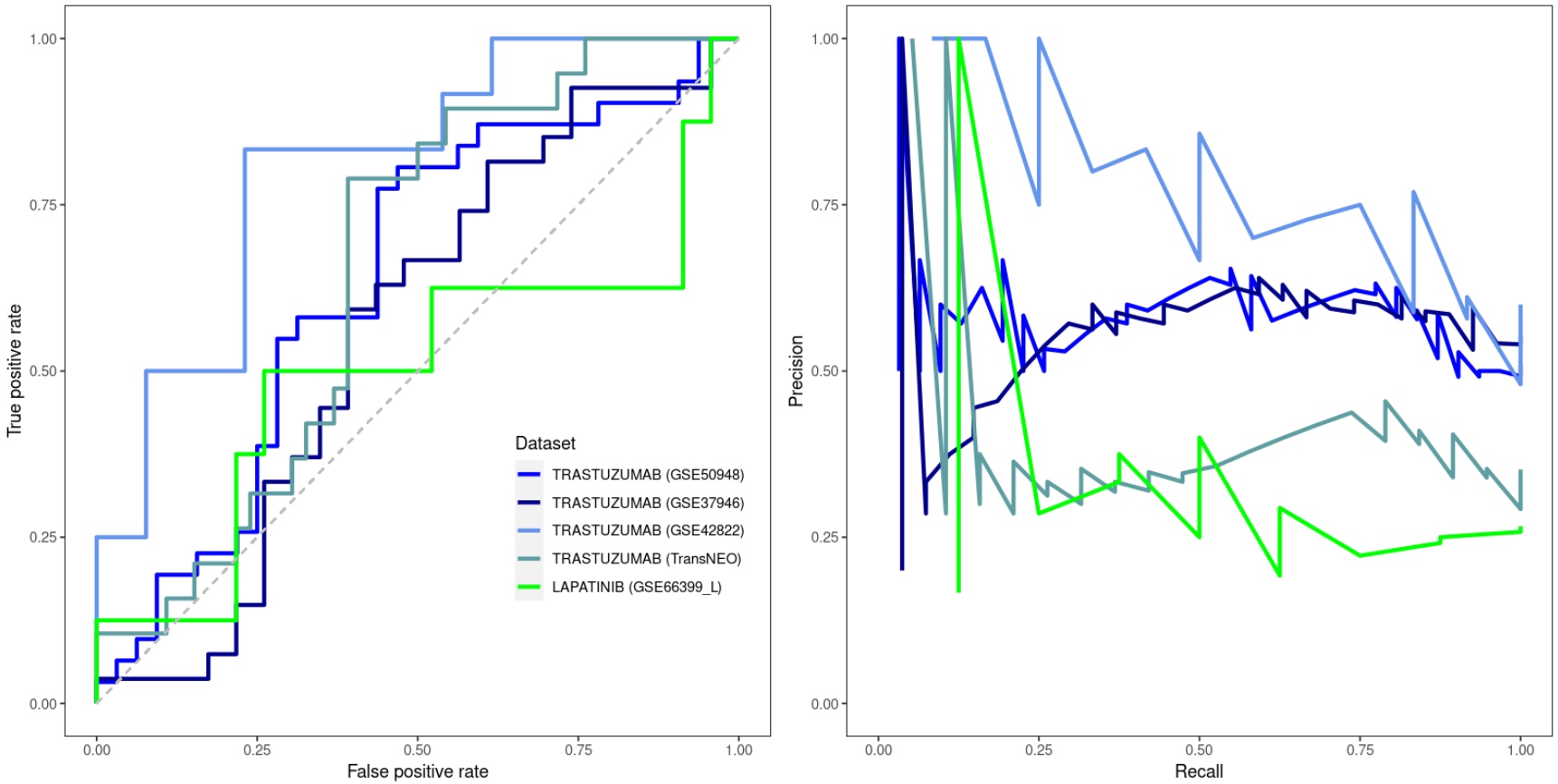
Table 2: ROC-AUC values and empirical p-values for BC-SELECT and its two score components (partner gene expression ratio and target gene expression) across 11 clinical trials of targeted therapy and immunotherapy. We report ROC-AUCs and describe the performance in the Monte Carlo "null distribution" with empirical p-values. ROC-AUC values are highlighted in bold and red if the ROC-AUC values fall within the top 2.5% of null AUC distributions. The reported ROC-AUC value is the higher of the analysis results between SL and SDL for the targeted therapy.
R: Responders. NR: Non-Responders. SL: Synthetic Lethality. SDL: Synthetic Dosage Lethality. SR: Synthetic Rescue.

Supplementary Figure 1

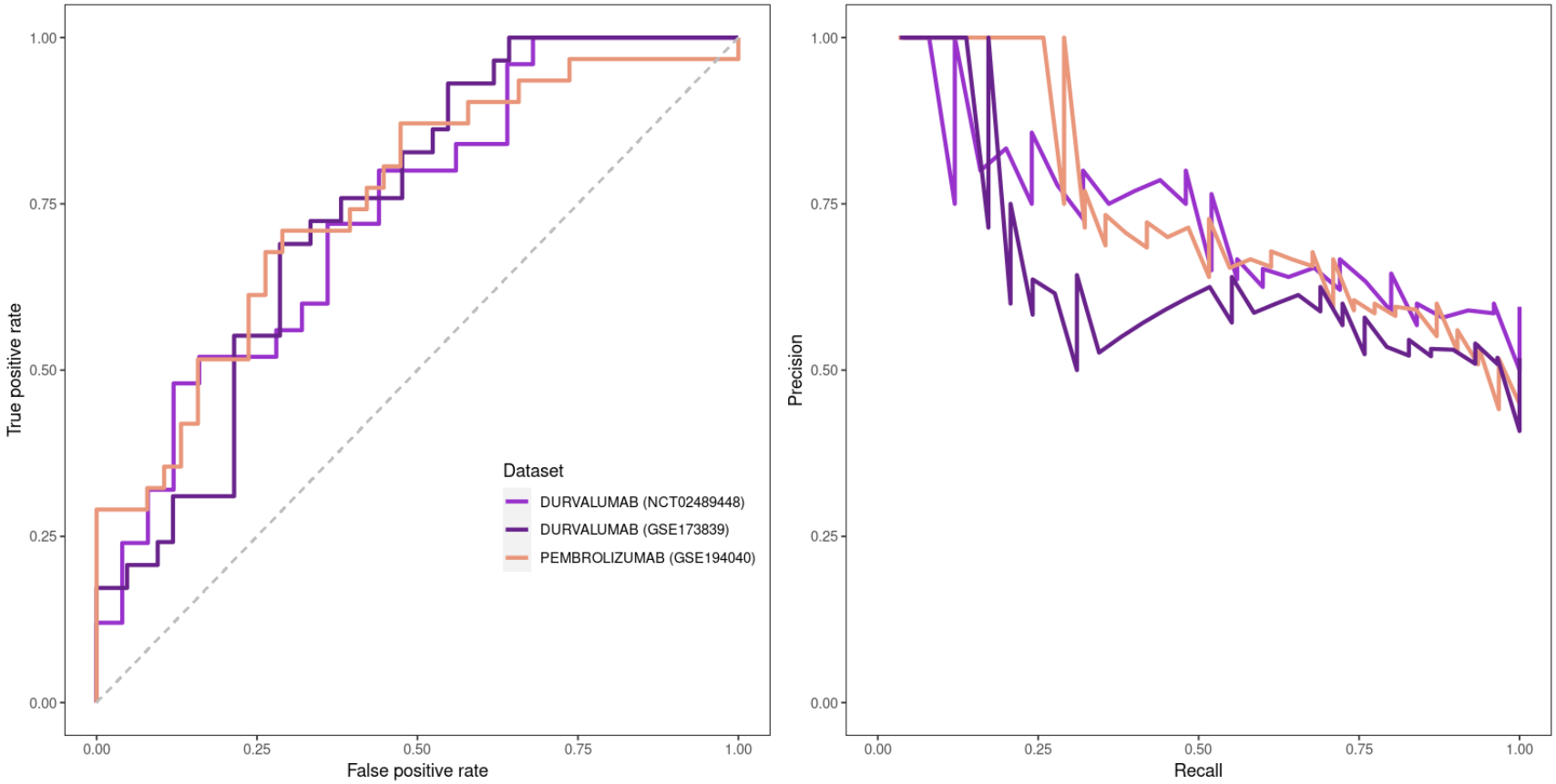
a.



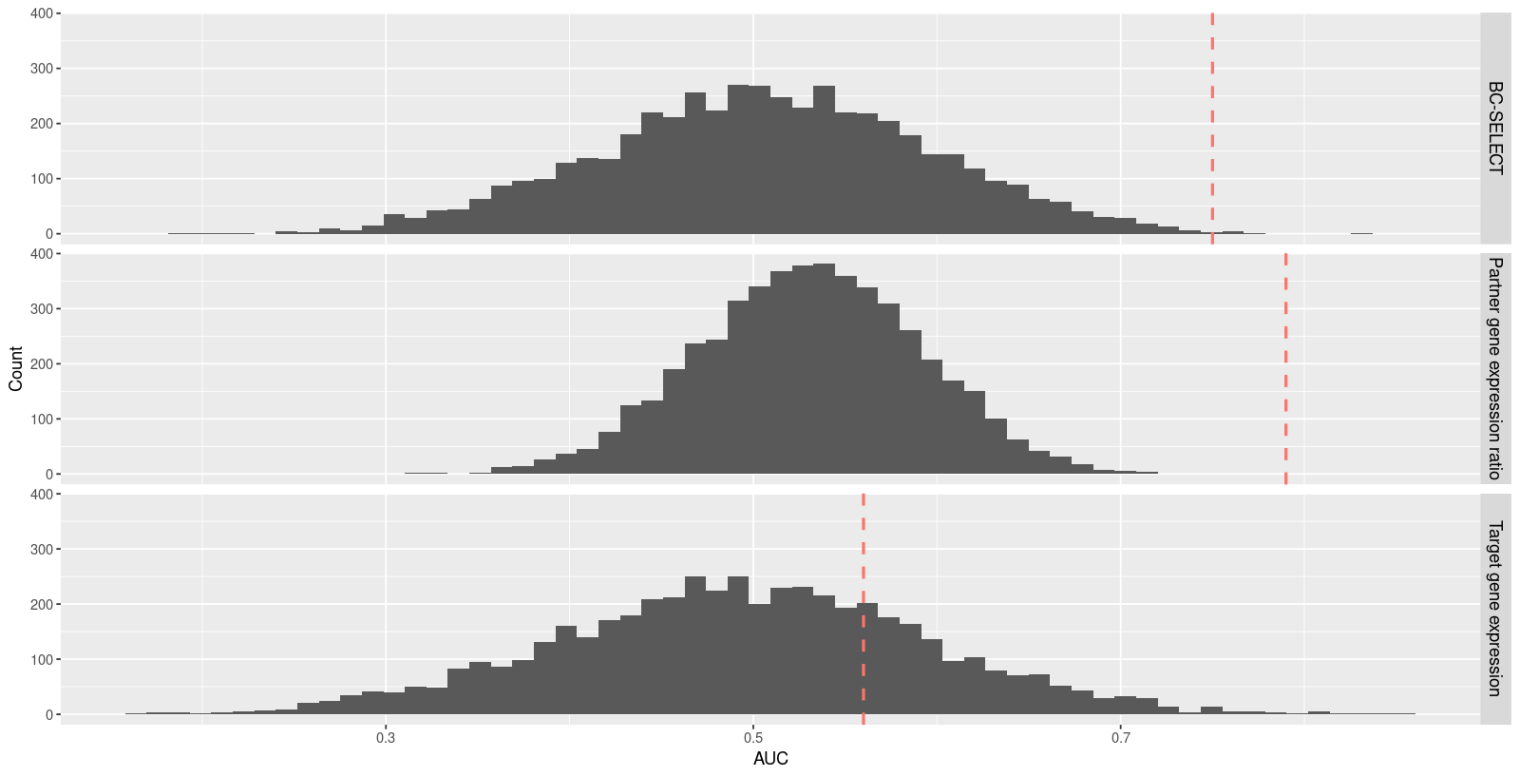
b.


True positive rate
False positive rate
0.00
0.25
0.50
0.75
1.00
Dataset
TRASTUZUMAB (GSE50948)
TRASTUZUMAB (GSE37946)
TRASTUZUMAB (GSE42822)
TRASTUZUMAB (TransNEO)
LAPATINIB (GSE66399_L)
Precision
Recall

c.



Supplementary Figure 1: ROC- and precision-recall curves across therapy types. The left panel shows the false positive rate plotted against the true positive rate for the ROC curve. The right panel shows the precision recall curve. **a. PARP inhibitors.** Veliparib-based trials are in red and olaparib is in orange. **b. HER2-directed therapy.** The right panel shows the precision recall curve. Trastuzumab-based trials are in blue and lapatinib is in green. **c. Immunotherapy.** The right panel shows the precision recall curve. Durvalumab-based trials are represented in purple, and pembrolizumab is shown in pink.



Supplementary Figure 2: Example of Monte Carlo-based "null distributions" for assessment of BC-SELECT performance of olaparib (GSE173839[57]). The X-axis depicts the distributions of 5000 ROC-AUCs calculated based on genes randomly selected to use for prediction. The null AUC distributions are generally centered around 0.5, indicative of random chance classification, while their varying right-tail thickness and the central tendency reflect differences reflective of the efficacy of prediction using gene expression. The three red dashed lines indicate the ROC-AUC values obtained by BC-SELECT, partner gene expression ratio, and target gene expression. Empirical p-values are calculated based on the proportion of the right tail area of the red dashed line and the histogram results of random selections. In this example, both BC-SELECT and partner gene expression ratio exhibit smaller empirical p-values, indicating superior performance of SL partner genes in drug response prediction compared to random guessing.