Application of Machine Learning to Predict the Biology Predicting response in the I-SPY 2 neoadjuvant breast cancer trial


UNBIASED PREDICTORS

An unbiased approach using all data yielded predictive power in 8 of 19 subgroups, including 5 with no predictive models from the first two approaches. Examples include HER2 ER, PIK3CA, HSP90 inhibitor; ALK inhibitors; and monoclonal antibody against CD47.

METHODOLOGY

In combination with clinical data, a three-pronged feature-selection approach was employed: (1) restricted to mechanism of action genes, (2) restricted to targeted therapies, and (3) restricted to targeted pathways for all 10 agents/combination plus ER and proliferation genes (n=330), all unbiased whole genome analyses. Results are summarized in Table 1. Models were considered predictive if AUROC > 0.75, Sensitivity ≥ 0.6 and Specificity ≥ 0.6 in cross validation and independent test sets.

BACKGROUND

Machine learning relies on algorithms that learn patterns in large, complex datasets to predict outcomes. The adaptive, neoadjuvant I-SPY 2 trial evaluates novel agents added to standard therapy, and incorporates measurement arms/hypothetical arms with no arms had strong predictive biomarkers. We leverage machine learning to explore the limitations of using only known mechanisms of action in predicting pCR, and the extent to which biological outcomes could inform response prediction in the first 10 arms of the trial.

I-SPY 2 is a multicenter, phase 2 trial using response-adaptive randomization within biomarker subtypes to evaluate a series of novel agents when added to standard neoadjuvant therapy for women with high-risk stage IIB breast (Figure 2). Within each patient subtype, participants are assigned to one of several investigational therapies or the control regimen: paclitaxel. Randomization probabilities are weighed by the probability of achieving a pCR within each subtype for each agent and adapt over the course of the trial to the primary endpoint is patients with a pCR (maternal luminal A, breast or nodal) at surgery.

Each treatment arm/receptor subtype subset with at least 20 patients (n=19) was evaluated independently with 25% of data held out as independent test sets. Log2 transformed expression was centered and scaled. We then used a 3-fold cross-validation technique with 10 repeats applying different resampling methods. Random Forest ensemble algorithms were compared with recursive feature elimination (Figure 3).

PREDICTION OF PATHOLOGIC COMPLETE RESPONSE

I-SPY 2 adaptive trial design

Table 1 summarizes the results of our analysis (red=Predictive model; blue=No-predictive model). Sensitivity, specificity, and the probability of predicting the pCR are shown for each treatment arm/hypothetical arm, as well as the probability of achieving a pCR within each subtype for each agent. The predictive model was developed using the primary endpoint of a pCR (maternal luminal A, breast or nodal) at surgery.

Results: The regimen predicted pCR in 2 of each of 14 subgroups across the three feature selection approaches: (1) restricted to mechanism of action genes; (2) restricted to targeted therapies; and (3) restricted to targeted pathways for all 10 agents/combination plus ER and proliferation genes. (3) An unbiased whole genome approach. Results are summarized in Table 2.

EMBEDDING MACHINERY OF ACTION

Expansion of the feature set to include genes associated with all mechanisms of action of each treatment arm yielded a model to predict good predictive models in 6 of 19 subgroups. Examples include DNA repair + immune gene plus predicting response to anti-HER2 to HER2+ cell lines (Figure 6).

CONCLUSIONS

Our results suggest that hypothesis driven analysis restricted to assumed mechanisms of action of the experimental agents may be insufficient, and that exploration of possible new biomarkers and interaction effects may be needed to understand the underlying biology of response to resistance.

ACKNOWLEDGMENTS:


iSPy2trial.org

This presentation is the intellectual property of the author/presenter. Contact them at: insert your email address here for permission to reprint and/or distribute.