Application of Machine Learning to elucidate the biology predicting response in the I-SPY 2 neoadjuvant breast cancer trial

R. W. Sayaman[1], D. M. Wolf[2], C. Yau[2], J. Wulfkuhle[3], E. Petricoin[3], L. Brown-Swigart[2], S. M. Asare[4], G. L. Hirst[2], L. Sit[2], N. O'Grady[2], D. Heditsian[5], **I-SPY 2 TRIAL Consortium**, L. J. Esserman[2], M. A. LaBarge[1], L. van 't Veer[2];

[1]City of Hope, Duarte, CA, [2]University of California, San Francisco, San Francisco, CA, [3]George Mason University, Fairfax, VA, [4]Quantum Leap Healthcare Collaborative, San Francisco, CA, [5]I-SPY 2 Advocacy Group, San Francisco, CA

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 identifies their most responsive subtype. While previously proposed genes/signatures reflecting an agent's mechanism of action predicted pathologic complete response (pCR) in some treatment arms/subtypes, not all 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 biology outside known drug action improves response prediction in the first 10 arms of the trial.

Our study involves 986 patients with pre-treatment gene expression and pCR data across 10 treatment arms including inhibitors of HER2: neratinib (N), pertuzumab (P), TDM1/pertuzumab (TDM1/P); AKT (MK-2206; M); IGF1R (ganitumab); HSP90 (ganetespib); PARP/DNA repair (veliparib/carboplatin; VC); ANG1/2 (AMG386); immune checkpoints (pembrolizumab; Pembro); and a shared control arm (Ctr) (Figure 1).

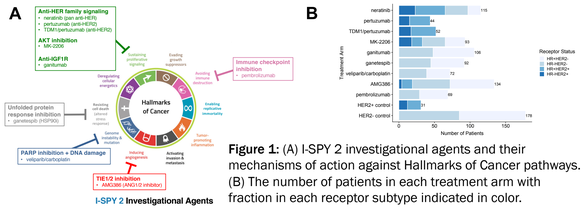


Figure 1: (A) I-SPY 2 investigational agents and their mechanisms of action against Hallmarks of Cancer pathways. (B) The number of patients in each treatment arm with fraction in each receptor subtype indicated in color.

I-SPY2's ADAPTIVE TRIAL DESIGN

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 II/III breast (Figure 2). Within each patient subtype, participants are assigned to one of several investigational therapies or the control regimen (4:1). Randomization probabilities are weighed by the probability of achieving a pCR within each subtype for each agent and adapts over the course of the trial. *The primary endpoint is pathologic complete response (pCR, no residual disease in breast or nodes) at surgery.*

The goal is to identify/graduate regimens that have ≥85% Bayesian predictive probability of success (statistical significance) in a 300-patient phase 3 neoadjuvant trial, defined by hormone-receptor (HR) & HER2 status & MammaPrint (MP).

Regimens may leave the trial for one of four reasons: Graduate, Drop for futility (< 10% probability of success), Drop for safety issues, or accruing maximum sample size (10%< probability of success <85%).

Figure 2: I-SPY 2 study schema and adaptive randomization based on probabilities of agents of achieving pCR within a given subtype.

METHODOLOGY

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

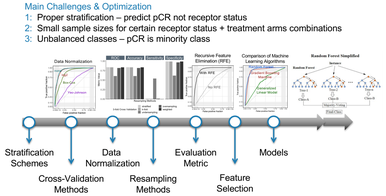


Figure 3: Schematic outlining main challenges, methododiical considerations and optimization steps. (Random Forest Simplified schematic from Wikimedia Commons)

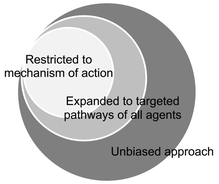


Figure 4: Three-pronged feature selection approach.

In combination with clinical data, a three-pronged feature-selection approach was employed: (1) restricted to mechanism of action genes: AKT/PI3K/HER (m=10 genes), IGF1 (m=11), HSP90 (m=88), DNA repair (m=79), TIE1/2 (m=11), and immune (m=61), as well as HER2 amplicon genes; (2) expanded to include targeted pathways for all 10 agents/combinations plus ESR1 and proliferation genes (m=339); (3) an unbiased whole genome approach (m=17,990) (Figure 4).

Models were considered predictive if AUROC ≥ 0.75, Sensitivity ≥ 0.6 and Specificity ≥ 0.6 in cross validation and independent test sets.

PREDICTION OF PATHOLOGIC COMPLETE RESPONSE

Table 1 summarizes the results of our analysis (red=Predictive model; blue=No predictive model; grey/NA=insufficient or no data). In total, we identified predictive biomarkers in 14 of 19 subgroups across the three feature selection approaches: (1) restricted to mechanism of action genes; (2) expanded to include targeted pathways for all 10 agents/combinations plus ESR1 and proliferation genes; (3) an unbiased whole genome approach.

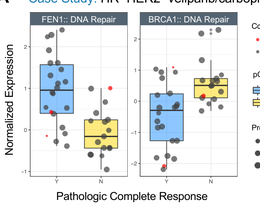
Table 1:

	HR+HER2-				HR-HER2-				HR+HER2+				HR-HER2+			
	Number of Patients	Results			Number of Patients	Results			Number of Patients	Results			Number of Patients	Results		
		Mechanism of Action	All Mechanisms of Action	Unbiased		Mechanism of Action	All Mechanisms of Action	Unbiased		Mechanism of Action	All Mechanisms of Action	Unbiased		Mechanism of Action	All Mechanisms of Action	Unbiased
neratinib (pan anti-HER)	34				32	No	No	Yes	42	No	Yes	Yes	23	No	No	Yes
pertuzumab (anti-HER2)	NA				NA				29	Yes	Yes	No	15			
TDM1/pertuzumab (anti-HER2)	NA				NA				35	No	No	Yes	17			
MK-2206 (AKT inhibitor)	28	No	No	No	32	No	No	Yes	16				18			
ganitumab (IGFR inhibitor)	58	No	Yes	No	48	No	No	No	NA				NA			
ganetespib (HSP90 inhibitor)	48	No	Yes	Yes	45	No	No	No	NA				NA			
veliparib/carboplatin (PARP inhibitor/DNA damage)	33	Yes	Yes	No	39	Yes	Yes	No	NA				NA			
AMG386 (ANG1/2 inhibitor)	62	No	Yes	Yes	53	No	No	No	15				6			
pembrolizumab (immune checkpoint inhibitor)	38	Yes	Yes	No	29	Yes	Yes	No	NA				NA			
HER2+ control	NA				NA				19				12			
HER2- control	94	No	No	Yes	84	No	No	No	NA				NA			

Insufficient sample size
No Predictive Model
Predictive Model

PREDICTORS WITHIN MECHANISM OF ACTION

Prediction of pCR using only genes reflecting the known mechanism of the drug succeeded in 5 subgroups, with DNA repair genes (Figure 5A) predicting VC response and immune genes (Figure 5B) predicting Pembro response in HR+HER2- and HR-HER2- subsets, and AKT/PI3K/HER + HER2 amplicon genes predicting pertuzumab response in HR+HER2+ patients.



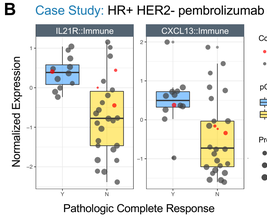


Figure 5. (A) Case Study HR-HER2- VC. DNA Repair genes, such as FEN1 and BRCA1, were predictive in 7 models. Two models consisted of 4 predictors each; 5 models of 9 predictors each. (B) Case Study HR+ HER2- Pembro. Immune genes, such as IL21R and CXCL13, were predictive in 2 models. One model consisted of 9 predictors; the ther of 49 predictors.

PREDICTORS ACROSS MECHANISMS OF ACTION

Expansion of the feature set to include genes associated with all mechanisms of action of all drugs proved sufficient to produce good predictive models in 8 of 19 subgroups. Examples include DNA repair + immune genes predicting response to ganitumab in HR+HER2- and to neratinib in HR+HER2+ (Figure 6).

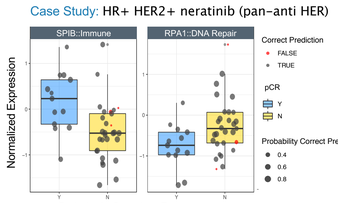
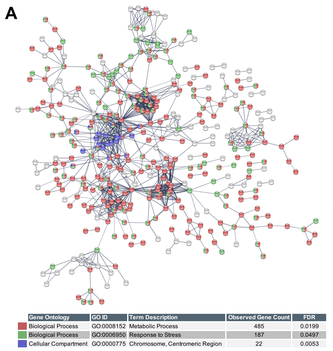


Figure 6. Case Study HR+ HER2+ neratinib. DNA Repair, such as RPA1, and immune genes, such as SPIB, which are outside the known mechanism of action of N were found to be predictive in 10 models. Five DNA repair and immune genes have predictive value in these models.

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 HR-HER2- neratinib predictors enriched for metabolic, cell division and membrane protein proteolytic processes; HR+HER2+ TDM1/P enriched for metabolic, stress response and cell cycle processes (Figure 7); and HR-HER2- MK-2206 predictors containing Ser/Thr kinases.



Gene Ontology	GO ID	Term Description	Observed Gene Count	FDR
Biological Process	GO:0008152	Metabolic Process	485	0.0199
Biological Process	GO:0006950	Response to Stress	187	0.0497
Cellular Compartment	GO:0000775	Chromosome, Centromeric Region	22	0.0053

Figure 7. Case Study HR+ HER2+ TDM1/P. (A) An unbiased approach discovers 902 ranked predictors, 867 are highly connected via protein-protein interaction (StringDB confidence > 0.7) and enrich for metabolic, stress response and cell cycle processes. (B-C) DAVID functional enrichment identifies a cluster of genes associated with cell cycle and mitosis that are upregulated (orange), e.g. NUP43, PTTG1, TUBB4B, ZWINT or downregulated (purple), e.g. RBBP8, in pCR vs. no pCR.

CONCLUSIONS

Our results suggests that hypothesis driven analysis restricted to assumed mechanisms of action of the experimental agents may be insufficient, and that exploration of possible off target effects may be needed to understand the underlying biology of response or resistance.

ACKNOWLEDGEMENTS:

With support from Quantum Leap Healthcare Collaborative, FNIH, NCI (Grant 28XS197 P-0518835), Safeway, an Albertsons Company, William K. Bowes, Jr. Foundation, Breast Cancer Research Foundation, UCSF, GMU, the Biomarkers Consortium, Salesforce, OpenClinica, Formedix, Natera, Hologic Inc., TGen, Illumina, CCS Associates, Berry Consultants, Breast Cancer Research – Atwater Trust, Stand up to Cancer, California Breast Cancer Research Program, and Give Breast Cancer the Boot, IQVIA, Genentech, Amgen, Pfizer, Merck, Seattle Genetics, Daiichi Sankyo, AstraZeneca, Dynavax Technologies, Puma Biotechnology, AbbVie, Madrigal Pharmaceuticals (formerly Synta Pharmaceuticals), Plexxikon, Regeneron and Agendia. Sincere thanks to our DSMB, Independent Agent Selection Committee, our patients, advocates and investigators.