Abstract number: 2112

RNA-based *ESR1* and *PGR* gene expression provides a real-time readout of hormone receptor status in patients with breast cancer

Lajos Pusztai[1], Saumya D. Sisoudiya[2], Rachel B. Keller-Evans[2], James J. Pao[2], Jeffrey S. Ross[2,3], Vidyalakshmi Sethunath[2], Richard S.P. Huang[2], Alexa B. Schrock[2], Lynette Poyser[2], Lucas Dennis[2], Garrett Frampton[2], Ericka M. Ebot[2], Ethan S. Sokol[2], Smruthy Sivakumar[2], Anand C. Annan[4]

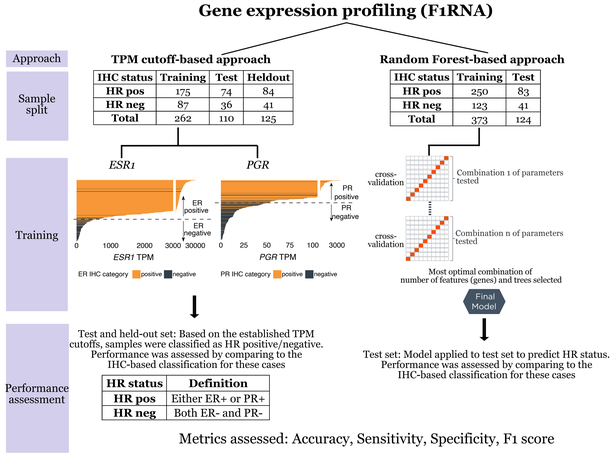
[1]Yale Cancer Center, New Haven, CT, USA [2]Foundation Medicine, Inc., Boston, MA, USA [3]Upstate Medical University, Syracuse, NY, USA [4]University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA

BACKGROUND

- Hormone receptor (HR) status is traditionally determined by IHC-based expression levels of the estrogen receptor (ER; encoded by *ESR1*) and progesterone receptor (PR; encoded by *PGR*) in tumor tissues and is used for clinical decision-making in breast cancer.
- With the increase in IHC-based predictive biomarkers, there is a need for development of tissue conservation strategies.
- Further, receptor subtype switching occurs in ~20% of breast cancer cases during progression, potentially impacting treatment decisions made based on IHC at initial diagnosis and thus requiring repeat testing.
- Multiplex RNA gene expression could serve as a screening or surrogate for protein expression by IHC especially when tissue availability is limited.
- Here we examined a cohort of patients with breast cancer profiled with a targeted RNA sequencing assay to evaluate two distinct approaches of utilizing gene expression for HR subtype inference- a TPM cutoff-based method and a machine learning-based method.

METHODS

- **Sequencing:** Study cohort comprised 497 patients with breast cancer who received RNA-based gene expression profiling using FoundationOne®RNA, a laboratory developed test by Foundation Medicine, Inc., capturing the expression of 1,517 genes as part of routine clinical care. IHC-based ER/PR annotation was extracted from pathology reports.
- **Algorithms:** Two algorithms were implemented to obtain HR status classification. One approach utilized *ESR1*/*PGR* expression (transcripts per million; TPM) level-based cutoffs that maximized accuracy in the training set**.** The second approach utilized a random forest-based machine learning algorithm with hyperparameter tuning through cross-validation in the training set. Performance of both approaches was assessed in independent test sets using the following metrics: accuracy, sensitivity, specificity, and F1 score.



RESULTS: OVERALL COHORT CHARACTERISTICS

Category	HR pos	HR neg	Total
Number of samples	333	164	497
Age median[IQR], y	65[54-73]	65[53-74]	65[53-73]
Biopsy site %[n]			
local	36% [121]	41% [68]	38% [189]
metastatic	44% [147]	40% [65]	43% [212]
lymph node	11% [38]	12% [19]	11% [57]
ambiguous	8% [27]	7% [12]	8% [39]
HRDsig positive %[n]	9% [30]	24% [38]	14% [68]
TMB median[IQR]	2.4[1.2-4.8]	3.6[1.8-4.8]	2.4[1.2-4.8]
MSS %[n]	96% [302]	97% [150]	96% [452]
Disease ontology %[n]			
Breast carcinoma (nos)	167	78	245
Breast invasive ductal carcinoma (IDC)	133	75	208
Breast invasive lobular carcinoma (ILC)	32	6	38

Table 1: Cohort summary
TMB:Tumor mutational burden; MSS: Microsatellite stable;
HRD: Homologous recombination deficiency

***ESR1*/*PGR* RNA expression was significantly higher in breast cancer samples with positive ER/PR IHC-based annotations compared to ER/PR negative tumors**

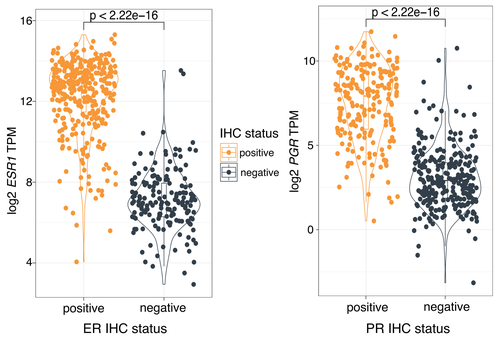


Figure 1: *ESR1* and *PGR* RNA expression by IHC status

RESULTS: HR STATUS PREDICTION MODEL PERFORMANCE

The Random Forest model had better specificity and sensitivity than the cutoff-based approach boosting the overall performance for HR status classification

Performance of TPM cutoff-based approach

Prediction \ Reference (IHC)	HR pos	HR neg
HR pos	79	5
HR neg	5	36

Accuracy	Sensitivity	Specificity	F1 score
92%	94%	88%	94%

Performance of Random Forest-based approach

Prediction \ Reference (IHC)	HR pos	HR neg
HR pos	79	3
HR neg	4	38

Accuracy	Sensitivity	Specificity	F1 score
94%	95%	93%	96%

Figure 2: Performance of the two models in the test sets

RESULTS: EXAMINATION OF DISCORDANT CASES

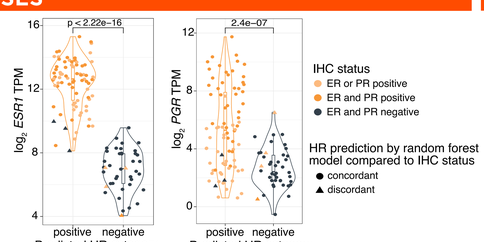


Figure 3: Random forest model-based HR status prediction vs IHC-based HR status of samples

- Of the 4 discordant samples predicted as HR negative by the Random Forest model, 2 were ER negative but PR positive based on IHC- showing that the model is largely driven by *ESR1* expression patterns
- The 3 discordant samples predicted as HR positive had relatively higher *ESR1* expression despite having IHC-based ER negative status. The %ER staining for these samples was between 0-4%

RESULTS: APPLICATION TO ER-LOW TUMORS

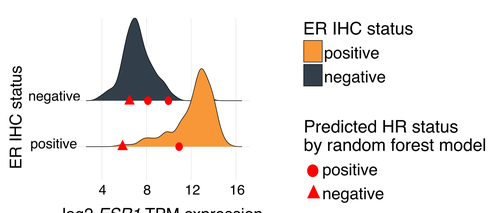


Figure 4: Samples with ER-low staining in the test set

- The test set had 5 samples that had low ER IHC staining (1-9%)
  - 2/5 were predicted as HR negative by the Random Forest model
    - Average log *ESR1* TPM expression of these samples was 6.2
  - 3/5 with positive HR predicted status had an average $\log_2$*ESR1* TPM of 9.7

CONCLUSION

- *ESR1* and *PGR* RNA expression is highly correlated with ER/PR IHC status, and gene expression-based prediction approaches from RNA expression profiling could potentially be utilized to infer real-time HR status with high accuracy.
- Random Forest based multi-gene model outperforms the TPM cutoff-based (*ESR1*/*PGR)* approach for predicting HR status.
- Utilizing a multiplex RNA expression methodology as a screen or surrogate for predictive biomarkers such as ER/PR will become increasingly important with more IHC-based predictive biomarkers on the horizon.
- Further validation is required to assess the utility of the random forest multi-gene expression model in clinical decision-making to support breast oncologists.

PLEASE DO NOT DISTRIBUTE WITHOUT THE AUTHOR'S OR FMI'S PERMISSION