Gene expression profiles accompanying phenotypic changes during non-malignant breast epithelial cells acini formation to explain MRI phenotypes

BIOARRAY Therapeutics www.bioarray.us

Marcia V. Fournier Ph.D.[1], Alan Derr[1], Alex Margulis Ph.D.[1], Kevin Reid[1], Sara Brumbaugh[2], Richard Anderson[3], W Fraser Symmans[4], ACRIN 6657 Trial investigators, Laura Esserman[5], and Nola Hylton[5].

Background

Magnetic resonance imaging (MRI) captures the three dimensional organization of tumors n the breast, defined as imaging phenotypes in the I-SPY 1 trial (CALGB 150007/150012; ACRIN 6657).

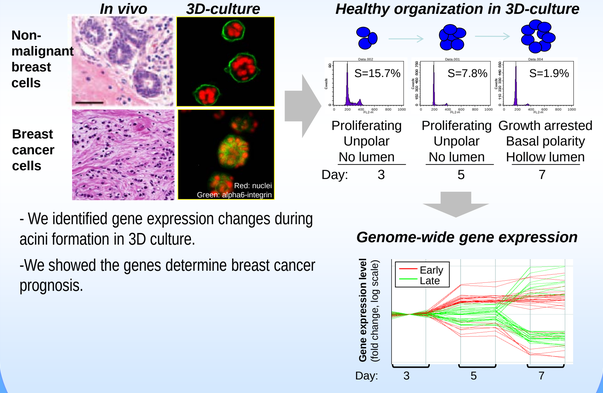
We developed a gene set based on the breast epithelial cells organization in three dimensional cultures. We investigated whether these organizational genes correspond to the imaging phenotypes.

MRI phenotypes have been shown to correspond to pathological complete response (pCR) to neoadjuvant chemotherapy, and are used to predict the ability to achieve breast conservation treatment (Mukhtar et al, Ann Surg Oncol, 20: 3823–3830, 2013).

We hypothesized that the molecular profile accompanying phenotypic changes occurring during the organization process of non-malignant acini may explain the molecular basis of MRI tumor phenotypes.

Approach

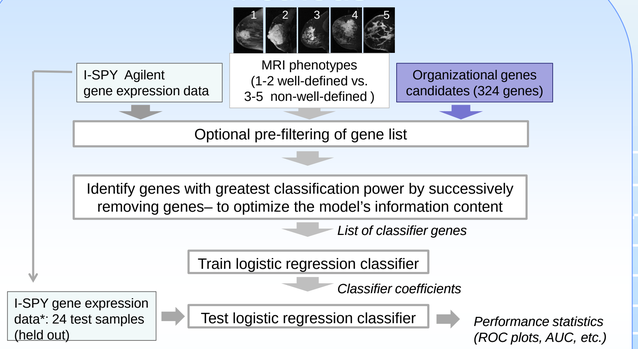
We have developed prediction models for MRI phenotypes and pCR based on expression profiles identified during the organization process of non-malignant breast epithelial cells in three-dimensional laminin-rich extracellular matrix.



- We identified gene expression changes during acini formation in 3D culture.
-We showed the genes determine breast cancer prognosis.

Figure 1. Approach to identify predictive genes is based on the healthy biology (Fournier, et al., 2006, Cancer Research 66:7095-7102).

Methods



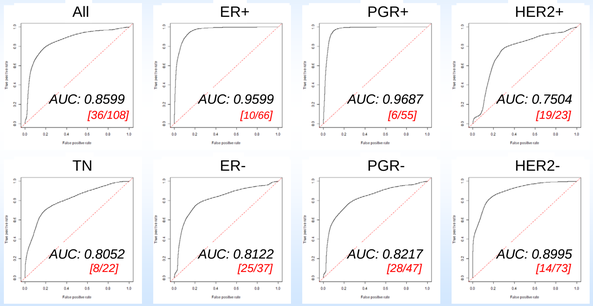
I-SPY datasets: GEO dataset GSE22226 Neoadjuvant breast cancer samples (pre-treatment) Agilent 2-color microarray: Channel 1 – reference RNA (Human Universal Ref., plus MCF7 and MC16C); Channel 2 – RNA extracted from test sample. Loess-normalized. 221 samples total, 147 samples with MRI phenotype annotation. 144 samples with pCR annotation, 36 pCR (25%), 108 non-pCR (75%). The distribution of phenotypes in I-SPY 1 was: 1-2 well-defined (49%), 3-5 non well-defined (51%).

Gene Selection: Two-step process: Select the best genes for the model. Using backward penalized stepwise regression, successively eliminate genes that are least useful to the classifier performance.

AUC (Area Under Curve): Using logistic regression-based classifiers R/Bioconductor functions glm, et al. Models are trained and tested 85/15 random partitioning of dataset: 85% used for training, 15% held out for independent test set. Using the AUC (area under the curve) of the ROC plot as a one-number accuracy metric for the predictors 0.5 = chance performance (50% accuracy); 1.0 = 100% accuracy.

Results

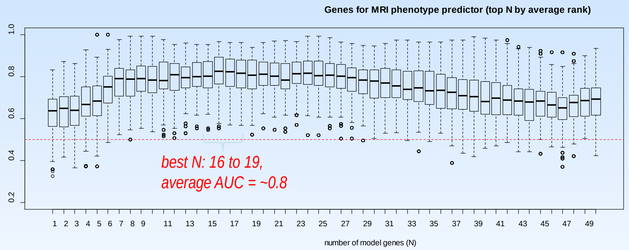
Figure 2. pCR predictors ROC plots by subtype



[pCR/non-pCR]

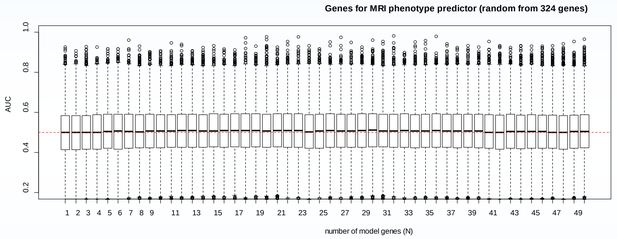
Results

Figure 3. MRI phenotype gene predictors



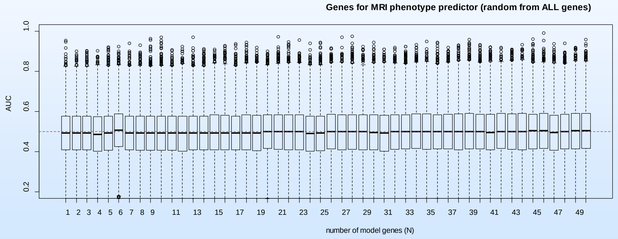
BIOARRAY Predictors' performance is excellent (AUC >0.8)

Procedure:
- Select top N genes by average rank
- Train/test on 10000 random partitioning (85%/15%)



Random Predictors' performance only slightly better than chance (i.e., AUC = 0.5)

Procedure:
- Select N genes randomly from 324 organizational genes
- Train/test on random partitioning (85%/15%)
- Repeat 10000 times

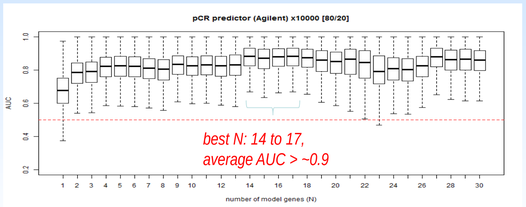


Random Predictors' performance the same as, or slightly worse than, chance (i.e., AUC = 0.5)

Procedure:
- Select N genes randomly from all genes
- Train/test on random partitioning (85%/15%)
- Repeat 10000 times

Results

Figure 4. Pathological complete response gene predictors



Procedure:
- Select top N genes by average rank
- End-point pCR
- Train/test on 10000 random partitioning (85%/15%)
- Random predictors performance slightly better than chance (random genes)

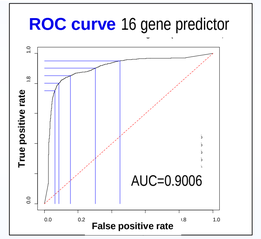
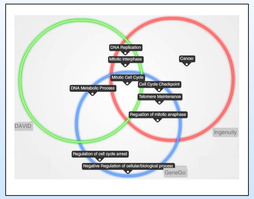


Figure 5. Prediction models included several genes that are known to regulate key cellular pathways such as cell division, metabolism, and migration using MetaCore pathway analysis. Cell cycle regulation was the most predominant function using gene ontology classification by pathway analysis software David, GeneGo, and Intuit.



Conclusions

Selection of the "best" genes from the 324 organizational gene set using our methodology results in models with very good prediction performance (avg. AUC ≈ 0.8) with much better performance than random selection.

The results suggest that BIOARRAY's biomarkers predicted MRI phenotypes, and can greatly improve prediction of pCR to standard chemotherapy across breast cancer subtypes.

Affiliations: 1) BIOARRAY Therapeutics Inc., Farmington, CT, 06032; 2) Ceres Analytics LLC, Redwood City, CA, 94063; 3) Symmetric Computing, Boston, MA, 02125; 4) University of Texas MD Anderson Cancer Center, Houston, TX, 77030; 5) UCSF Helen Diller Family Comprehensive Cancer Center, San Francisco, CA, 94143. **Acknowledgements:** Connecticut Innovations Inc., Avon Foundation.