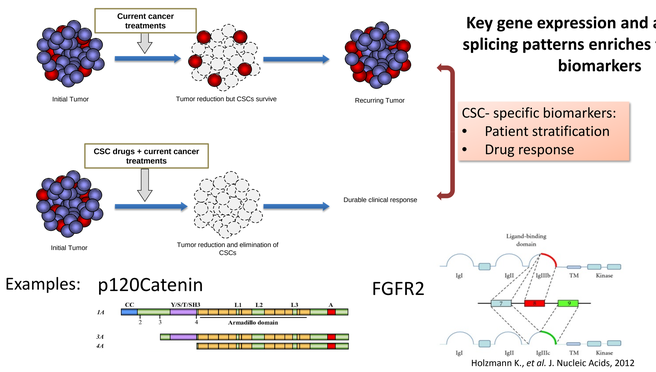
Gene Expression and Alternative Splicing Signatures Discriminate Breast Cancer Stem Cells from Fibroblasts

David T. Weaver, Irina M. Shapiro, Alan G. Derr, Daniel W. Paterson, Jonathan A. Pachter
Verastem, Inc., Cambridge, MA

1057

INTRODUCTION

Tumors frequently contain cancer stem cells (CSCs) or tumor-initiating subpopulations, with an ability to self-renew and regenerate all cell types within the tumor. Basal-like breast cancers exhibit features of CSCs, including expression of surface markers, even though these cells are rare. Given the role of CSCs in the recurrence and spread of cancer, there is an urgent need to develop new therapeutic agents that target CSCs. Development of CSC-targeted drugs will be greatly facilitated by biomarkers that can identify CSCs to aid in patient selection and determination of drug response. Defining the CSCs in tumors is complicated by the high mesenchymal nature of fibroblasts. Analysis of gene expression and alternative splicing patterns in CSCs that are not observed in fibroblasts may provide valuable new CSC-specific markers.



METHODS

Gene expression and alternative splicing whole transcriptome microarray strategies were used to identify selected differentially expressed genes and exons and between 10 Basal human breast cancer cell lines and a combination of 12 Luminal and 3 fibroblast cell lines. Gene signatures for Tumor Initiating (TI) and Epthelial-Mesenchymal Transition (EMT) were applied to resolve the patterns of expression.

Q-PCR analysis was conducted to validate candidate CSC genes from the above discovery method. Differential expression between Basal, Luminal, and Fibroblast cell lines, tumorsphere culture and human tumor xenograft experiments were evaluated.

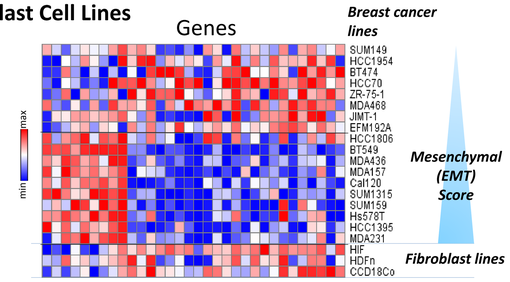
RNA microarrays from n=178 TNBC patients (GEO GSE25066) were evaluated with Time to Recurrence (0-7 years) models from the Mesenchymal-High, Epithelial-Low, Fibroblast-Low genes in this study. The accuracy of the model was assessed by ROC curves.

In a separate discovery effort, alternative splicing discriminators were assembled according to the FlowChart:

RESULTS

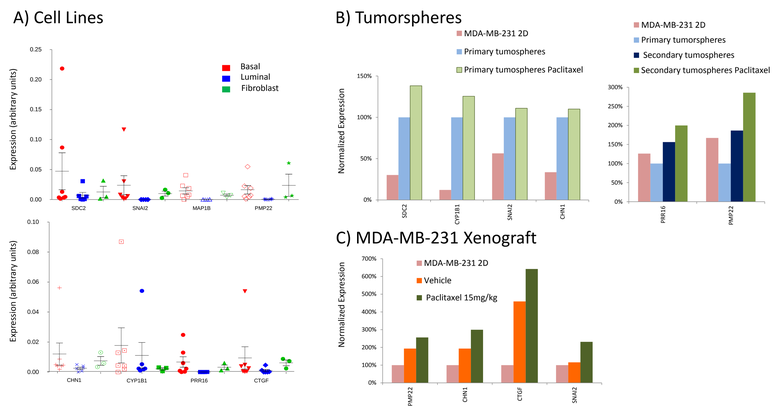
Expression levels of 11 genes were higher and 24 genes were lower in the Basal cell lines versus Luminal or fibroblastic cell lines. Comparison of Basal cell lines to the Luminal/Fibroblast cell lines identified 36 cassette exons that were included, and 26 that were excluded in Basal cell lines. Also, 19 genes were upregulated in Basal cell lines compared to the other groups as detected by Q-PCR. Interestingly, the 19-multigene model defined the Triple Negative Breast Cancer patients that were Likely to Recur under standard chemotherapy with a p = 1.90e-03 and AUC 0.723.

Fig 1: Differential gene expression between Basal, Luminal, and Fibroblast Cell Lines



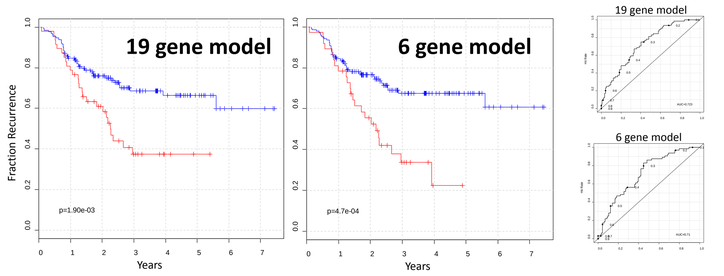
Relative normalized gene expression is shown for a group of genes having Basal cell line-high, Luminal cell line-low, and Fibroblast cell line-low relative expression. Clustering of breast cancer cell lines is conducted to reveal the differential expression between cell lines that are highest to lowest in Mesenchymal type, based on a Gene Set Enrichment Analysis EMT score.

Fig 2: RNA expression of Genes in Breast Cancer Cell Lines, Tumorspheres, and Tumor Xenograft.



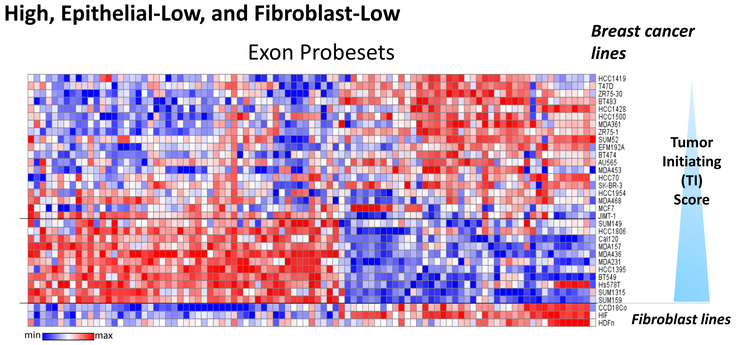
Q-PCR determinations of RNAs isolated from various sources. A) Basal, Luminal, or Fibroblast cell lines; B) Primary tumorspheres with or without in vitro treatment with Paclitaxel. C) MDA_MB_231 Breast Cancer Xenograft comparing RNA expression after vehicle and paclitaxel treatment. Relative expression values are shown.

Fig 3: Gene Expression Models for Triple Negative Breast Cancer Recurrence



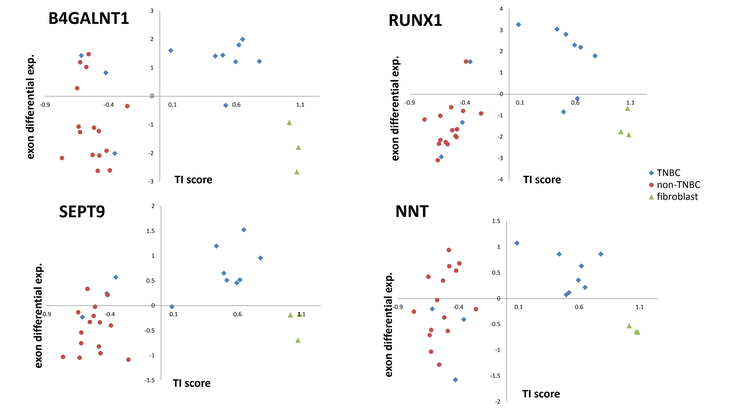
Cancer specimens from Triple Negative Breast Cancer patients (N=178) were examined for Gene Signature RNA expression based on candidate gene lists from the above analysis. 19 genes from the full set of Basal-High, Luminal-Low, Fibroblast-Low and a 6 gene subset of the same genes were modeled. Shown is the Time to Recurrence for the fraction of patients. On the right, are Receiver Operator Curves (ROC), and Area Under the Curve (AUC) values for these models.

Fig 4: Alternative Splicing in Basal Cell Lines that are Mesenchymal-High, Epithelial-Low, and Fibroblast-Low



Relative normalized expression for exon probesets is shown for a group of exons that are differentially expressed between breast cancer cell lines. The cell lines are ordered based on Tumor Initiating (TI) Score, a gene rank algorithm derived from the Creighton et al. (2009) gene signature. High TI score is on the bottom compared with Low TI score on the top.

Fig 5: Differential Gene Isoform Levels between TNBC, other Breast Cancer, and Fibroblasts compared with Tumor Initiating Score



Differentially expressed exons for 4 genes are shown relative to a Tumor Initiating (TI) gene rank algorithm. High positive TI score is indicative of cells that are high in tumor initiating capability. High level of exon differential expression is noted as positive values.

SUMMARY

- Differentially expressed genes and alternatively spliced exons were discovered as biomarkers of cancer stem cells in breast cancer cell lines, distinct from fibroblasts.
- Gene and exon marker sets that distinguish CSCs versus fibroblasts and may be instructive in identifying patients that recur early in Triple Negative Breast Cancer.
- The CSC-associated RNA signatures identified here will be further refined to develop new CSC-specific diagnostic biomarkers to stratify breast cancer patients and monitor response to novel CSC-targeted therapies.

Verastem
novel drugs targeting cancer stem cells

Verastem, Inc. | 215 First Street | Suite 440 | Cambridge | Massachusetts | 02142