Development of Prognostic Gene Panels: Subtype Prediction & Risk Stratification in Breast Cancer

Irshad Mohammed, Nidhi Singh*
Pattern Computer, Inc., www.patterncomputer.com

INTRODUCTION

- Breast cancer is comprised of different entities, each being associated with different outcomes and therapeutic approaches.
- Based on gene-expression profiling, five major molecular subtypes exist namely, Luminal A, Luminal B, Basal, Her2 and Normal-like.[1]
- Recent studies have shown that even though the five intrinsic subtypes share clinical and biological features, there is substantial variation within each group in terms of survival.[2]
- The aim of this study was to develop better and more accurate and robust multivariate prediction models that are capable of simultaneously classifying tumors by their molecular subtypes while accurately stratifying them into low-risk and high-risk disease-states to inform treatment decisions in breast cancer.
- These results indicate that a prognostic panel with higher resolution in risk stratification may lead to improved therapies in precision medicine for patients with breast cancer.

METHOD

- Using feature-selection engine, high-dimensional genomic data (gene expression data from METABRIC cohort of 2000 breast cancer samples) was reduced from around 20,000 features to the order of 10s of genes.
- Multiple gene panels were derived using our proprietary machine learning tools, which enabled the identification of the top-weighted genes that, together, reproducibly identify subtype and survival, Fig. 1. This was followed by retraining the calibration engine with gene panels with varying numbers of genes to enhance predictive power.
- The overall accuracy for the calibrated model (**Pattern BC38**), as shown in Fig. 2 was found to be ~90% that was improved by repeat testing of tumor sub-samples under a Bayesian model to 99%.
- We also developed gene panels as survival predictors - defining risk groups as patients who died due to disease in less than 5 years of diagnosis - as high-risk vs patients (living/dead) who lived beyond 16 years – as low risk, Fig. 3.
- The top 6 genes account for 95.5% of the variability of the Pattern BC38, prompting us to study a reduced six-gene panel, **Pattern BC06** that can adequately classify subtypes as well as survival risk , as shown in Fig. 4.
- The performance of our panel was assessed on external, independent breast cancer datasets.
- It was found that the simplified gene panel had an overall **prediction accuracy of ~86%** for test samples, which we project will obtain >99% accuracy after testing in biological quadruplicate.

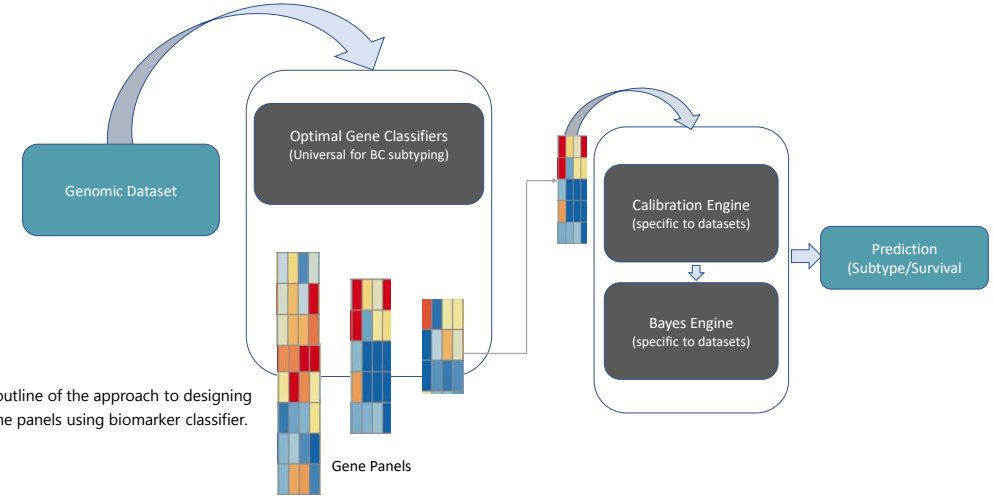


Figure 1. An outline of the approach to designing classifying gene panels using biomarker classifier.

RESULTS

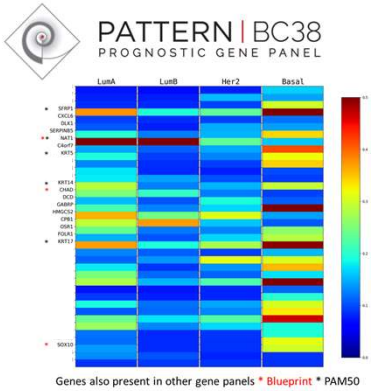


Figure 2a. The **Pattern BC38 gene panel** for breast cancer subtype and survival classification. The bar next to it shows expression levels from low-blue to high-red.

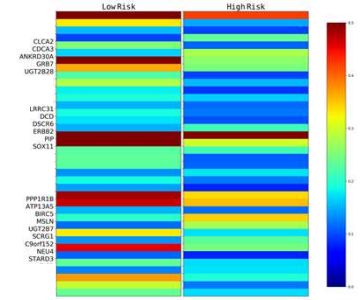
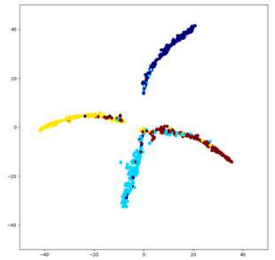
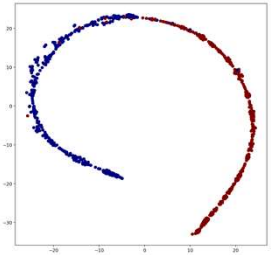


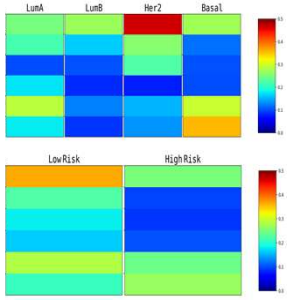
Figure 3a. The gene panel for breast cancer survival classification.



2b. A 2D representation of breast cancer subtypes generated using t-SNE dimensional reduction technique.



3b. A 2D representation of risk classification in breast cancer generated using t-SNE dimensional reduction technique.



**Redacted genes represent proprietary PCI content.

Figure 4. The Pattern BC06 gene panel for breast cancer subtype and survival classification.

CONCLUSIONS

- We have developed a robust and cost-efficient biomarker with a set of only six genes (Pattern BC06) that can predict both survival and subtypes with an accuracy of 86%.
- Our biomarker panel will allow an improved understanding of clinical implications of molecular subtypes and risk prediction that could help clinicians guide precision medicine, tailoring medical treatment to patients and their tumor characteristics.

REFERENCES

1. Perou CM, Sørlie T, et al. Molecular portraits of human breast tumours. Nature. 2000 Aug 17;406(6797):747-52.
2. Sørlie T, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001 Sep 11;98(19):10869-74.