Unveiling Survival Signatures: Assessing Contributions of Transcriptional Outliers on Breast Cancer Survival

JAMES R. EVANS†, AMAAN M. JOGIA-SATTAR†, Jee Y. Han2,3,4,5, Zhuyu Qiu2,3,4,5, Paul C. Boutros2,3,4

Introduction | Abstract

Despite their potential as prognostic biomarkers and therapeutic targets, outlier genes with unique expression patterns in cancer have not been comprehensively characterized in the context of patient survival. The integration of outlier gene signatures with clinical measurements enables more nuanced profiling of tumor-specific heterogeneity. Leveraging TCGA RNA-Seq and METABRIC microarray data, we constructed univariate and multivariate Cox proportional hazards models to investigate the impact of outlier genes on patient survival. Our findings shed light on the role of outlier genes as proxies for patient survival, demonstrating their relevance in characterizing breast cancer survival patterns. Exploration of clinical variables in association with outlier genes and their molecular mechanisms can unveil avenues for therapeutic interventions targeting tumor-specific transcriptional variations. Future endeavors include an extension of this methodology across cancer types and advancing pattern identification in pan-cancer outlier contributions for clinical and molecular data.

Methods | Schematic Workflow

Methods / Schematic Workflow

TCGA-BRCA Dataset:
- RNA-seq data from n = 1085 breast cancer patients, with 1187 identified outlier genes
- RNA abundance levels represented as Fragments Per Kilobase Million (FPKM)

METABRIC Dataset:
- Microarray data from n = 1991 breast cancer patients, with 4782 identified outlier genes
- RNA abundance levels represented as probe intensity values

Results | Discussion

TCGA-BRCA & METABRIC Outlier Landscape

FIGURE 1. TCGA-BRCA Outlier Heat Map. Clustering of outliers by gene and patient

Survival Analysis: Outlier Gene Count and Molecular Subtype

FIGURE 5. Kaplan-Meier Survival Curves and Cox Proportional Hazards Models Stratified by Outlier Count

Clinical Variable Correlation

FIGURE 3. Contingency Matrices Displaying Outlier Count by Molecular Subtype

Survival Analysis: Outlier Gene Contributions

FIGURE 7. FDR-Adjusted P-Values from Gene-Wise Univariate Cox Proportional Hazards Modeling

Conclusion

- ER and PR status became increasingly negative with increasing outlier gene count.
- HER2 status became increasingly positive with increasing outlier gene count.
- METABRIC displayed reduced survival probability with increasing outlier gene count early post-diagnosis. TCGA had limited follow-up, requiring further investigation.
- Both diagnosis age and radiation therapy showed significant associations with survival in METABRIC covariate-adjusted Cox models, while diagnosis age was significantly associated in TCGA.
- METABRIC revealed significant outlier genes post-FDR correction. Limited significant outliers in TCGA; longer follow-up may clarify.
- Future endeavors include exploration of outlier genes’ survival patterns in other cancer types for pan-cancer insight generation, as well as replication of survival analyses in cohorts with extended patient follow-up.

References


Acknowledgements | Funding

We would like to thank B.I.G. Summer with support from the R25-NIDCR (BIG DOC) for funding our research experience.

Author Affiliations

1 BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
2 Department of Human Genomics, University of California, Los Angeles, USA
3 Department of Urology, University of California, Los Angeles, USA
4 Jonsson Comprehensive Cancer Center, University of California, Los Angeles, USA
5 Institute for Precision Health, University of California, Los Angeles, USA
6 Department of Medical Biophysics, University of Toronto, Toronto, Canada

* equal contributions