A Novel miRNA Expression-Based Classification of Breast Cancer Tumours

Dongkyu Lee, Wayne Fan, Tanmay Thakral

Introduction

Breast cancer was the leading cause of cancer death among women aged 20-59 years in the United States in 2017 (1). Early diagnosis is often crucial for better prognosis.

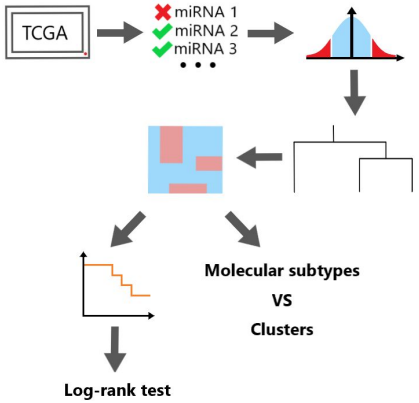
Recent studies suggest using differentially expressed miRNAs, called miRNA signatures, as both diagnostic and prognostic tools in the context of breast cancer (2).

miRNAs are small non-coding RNA strands that regulate gene expression via gene silencing and post-translational modification. As such, they have significant influence over cellular activity, including tumour cell activity.

While previous studies have investigated differential miRNA expression between molecular subtypes, a well established breast cancer tumour classification system, fewer studies have focused on alternative methods of classifying breast cancer tumours (3).

Research Objective: To use miRNA expression profiles from breast cancer patients to generate a novel breast cancer classification system.

Methods



- Import patient clinical and miRNA expression data from the TCGA database to a Python dataframe
- Filter out outliers and data points with statistically low variability
- Normalize values in the filtered dataset
- Unsupervised hierarchical clustering to generate a dendrogram
- Visualize clusters by generating a heatmap
- Chi-squared test to compare molecular subtypes with clusters
- Perform survival analysis on clusters
- Raw data and code are available through QR code at the top right

Results

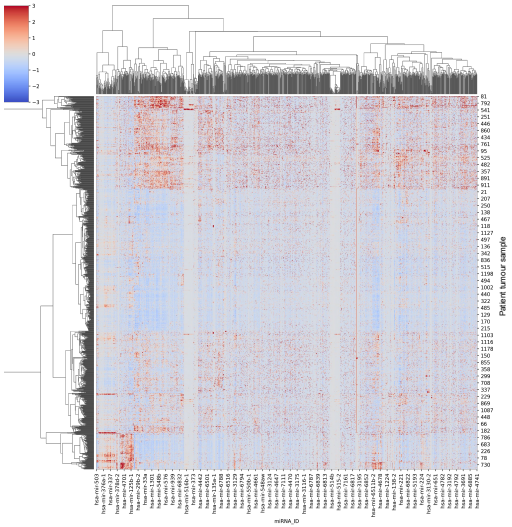


Figure 1: Heat map shows clustering pattern
A total of 1100 miRNAs and 1203 tumour samples are represented. Red represents high expression, while blue represents low expression. Dendrograms represent the relatedness of miRNAs (x-axis) or patient tumour samples (y-axis).

Cluster	Total Count	Subtype				
		HER2-Enriched	Luminal A	Luminal B	Triple-Negative	Unknown
0	119	3	47	5	11	53
1	259	11	88	3	30	127
2	785	25	266	13	78	403
3	39	2	7	2	5	23

Table 1: Molecular subtyping of tumours
The molecular subtype of each patient tumour sample in each cluster is determined. Unknown indicates ambiguous or indeterminate data.

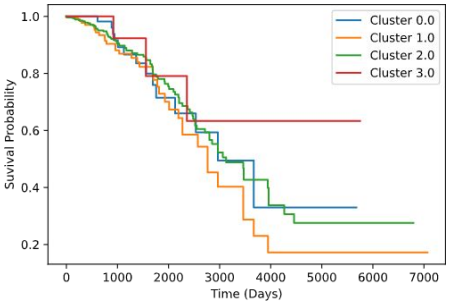


Figure 2: Survival curve of tumour clusters
The decreasing survival rate over time for each cluster is plotted on a Kaplan meier survival plot.

Discussion

Tumour clusters are not indicative of molecular subtypes

- Chi-squared test yields a p-value of 0.24, which is greater than the significance level (0.05).
- Therefore, miRNA expression patterns of these 4 tumour clusters suggest a novel classification system.
- Previously studied individual miRNAs are statistically reliable biomarkers for molecular subtypes (4).
- miRNAs may be less reliable for indicating molecular subtypes when aggregated into large groups.

Tumour clusters have similar survival outcomes

- Log rank tests between each unique cluster pair combination yielded p-values greater than the significance level (0.05).
- Prognosis-contributing miRNAs have previously been identified to be subtype-specific (5).
- Since all molecular subtypes are present in each tumour cluster, similar survival outcomes across tumour clusters could be due to the effect of multiple prognosis-contributing miRNAs.

Limitations

- The subtyping method used in this research project is less accurate compared to PAM50 subtyping (6).
- The dataset for this research project only includes conserved miRNA across the body. Thus, highly specific breast cancer related miRNA biomarkers may be overlooked in our analysis

Future Directions

- Reduce noise in our analysis (minimize the number of miRNAs/ patients)
- Test the correlation between clusters and other categories of patient outcome (e.g. recurrence, metastasis, etc.)
- Analyze different datasets

Acknowledgements

We would like to thank our mentor, Eric Lee, for sharing his expertise and providing critical feedback during the course of this project.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA: A Cancer Journal for Clinicians. 2020;70(1):7–30.
2. Lerebours F, Cizeron-Clairac G, Susini A, Vacher S, Mouret-Fourme E, Belichard C, et al. miRNA expression profiling of inflammatory breast cancer identifies a 5-miRNA signature predictive of breast tumor aggressiveness. International Journal of Cancer. 2013;133(7):1614–23.
3. Haakensen VD, Nygaard V, Greger L, Aure MR, Fromm B, Bukholm IRK, et al. Subtype-specific micro-RNA expression signatures in breast cancer progression. International Journal of Cancer. 2016;139(5):1117–28.
4. Søkilde R, Persson H, Ehinger A, Pirona AC, Fernö M, Hegardt C, et al. Refinement of breast cancer molecular classification by miRNA expression profiles. BMC Genomics. 2019 Jun 17;20(1):503.
5. He K, Li W-X, Guan D, Gong M, Ye S, Fang Z, et al. Regulatory network reconstruction of five essential microRNAs for survival analysis in breast cancer by integrating miRNA and mRNA expression datasets. Funct Integr Genomics. 2019 Jul 1;19(4):645–58.
6. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. JCO. 2009 Feb 9;27(8):1160–7.