An isomiR expression panel based novel breast cancer classification via using improved mutual information*

Chaowang Lan, Hui Peng, Eileen M.McGowan*, Gyorgy Hutvagner* and Jinyan Li*

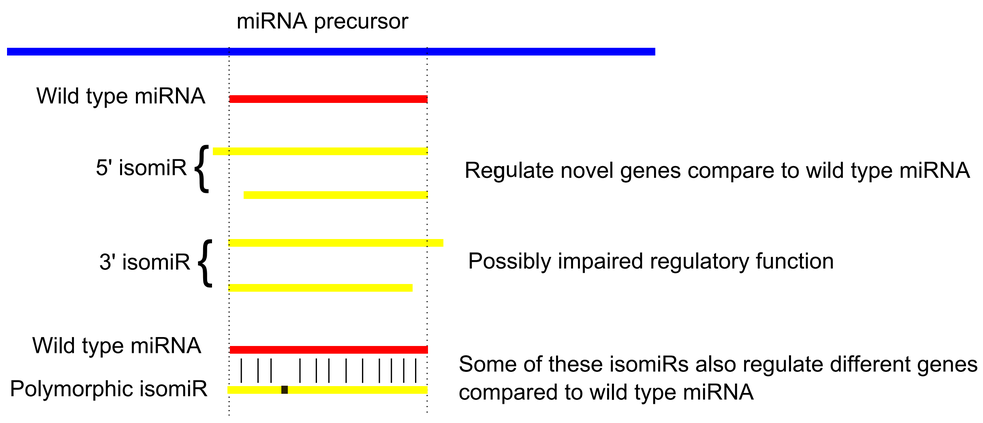
University of Technology, Sydney

E-mail:Chaowang.Lan@student.uts.edu.au

BACKGROUND

- Understanding the mechanisms of breast cancer subtyping is clinically useful with respect to prognosis, prediction, and informed therapeutic choices.
- IsomiRs are isoforms of miRNAs have been successfully used to distinguish various cancer types.
- Biomarker isomiRs for identifying different breast cancer subtypes has not been investigated.

The three main categories of isomiRs



Materials

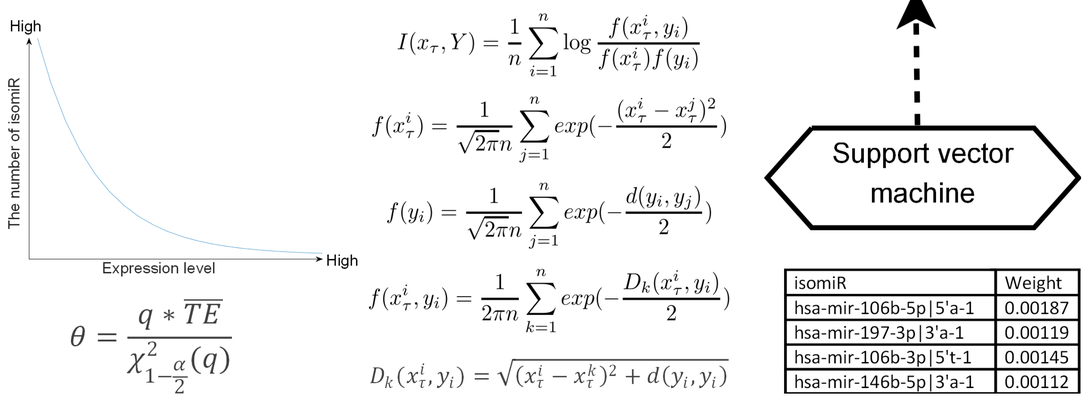
The expression profiles of isomiRs in breast cancer patients can be downloaded from TCGA GDC data portal website.

Breast cancer subtype reclassification for isomiR identification.

Subtype name	ERα+HER2-	ERα-HER2+	ERα+HER2+	Triple negative
Number of patient	472	31	76	119

Method

IsomiR biomarker subtyping methodology



$$I(x_\tau, Y) = \frac{1}{n}\sum_{i=1}^{n} \log \frac{f(x_\tau^i, y_i)}{f(x_\tau^i) f(y_i)}$$

$$f(x_\tau^i) = \frac{1}{\sqrt{2\pi}n}\sum_{j=1}^{n} exp(-\frac{(x_\tau^i - x_\tau^j)^2}{2})$$

$$f(y_i) = \frac{1}{\sqrt{2\pi}n}\sum_{j=1}^{n} exp(-\frac{d(y_i, y_j)}{2})$$

$$f(x_\tau^i, y_i) = \frac{1}{2\pi n}\sum_{k=1}^{n} exp(-\frac{D_k(x_\tau^i, y_i)}{2})$$

$$D_k(x_\tau^i, y_i) = \sqrt{(x_\tau^i - x_\tau^k)^2 + d(y_i, y_i)}$$

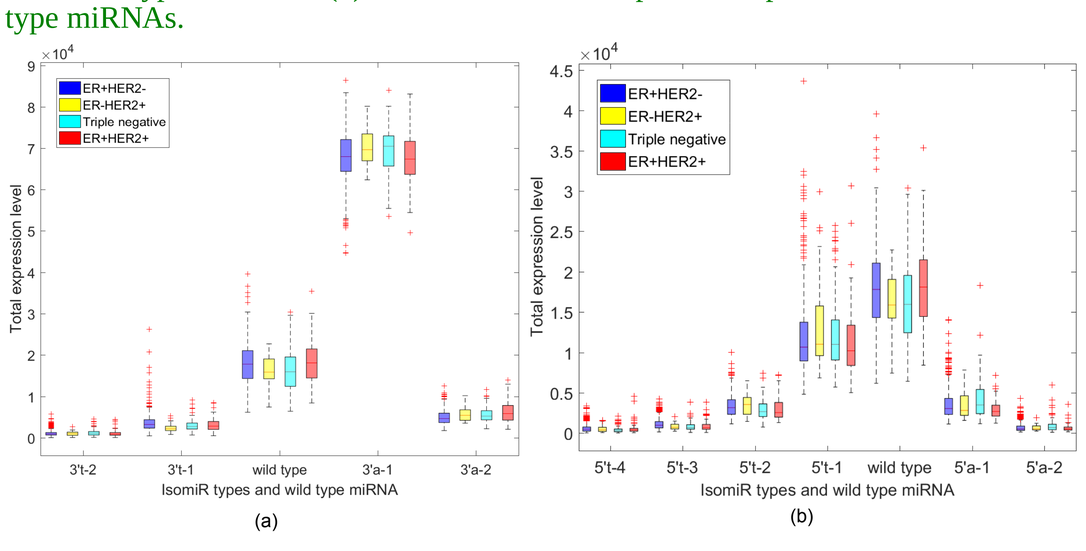
$$\theta = \frac{q * \overline{TE}}{\chi^2_{1-\frac{\alpha}{2}}(q)}$$

isomiR	Weight
hsa-mir-106b-5p|5'a-1	0.00187
hsa-mir-197-3p|3'a-1	0.00119
hsa-mir-106b-3p|5't-1	0.00145
hsa-mir-146b-5p|3'a-1	0.00112

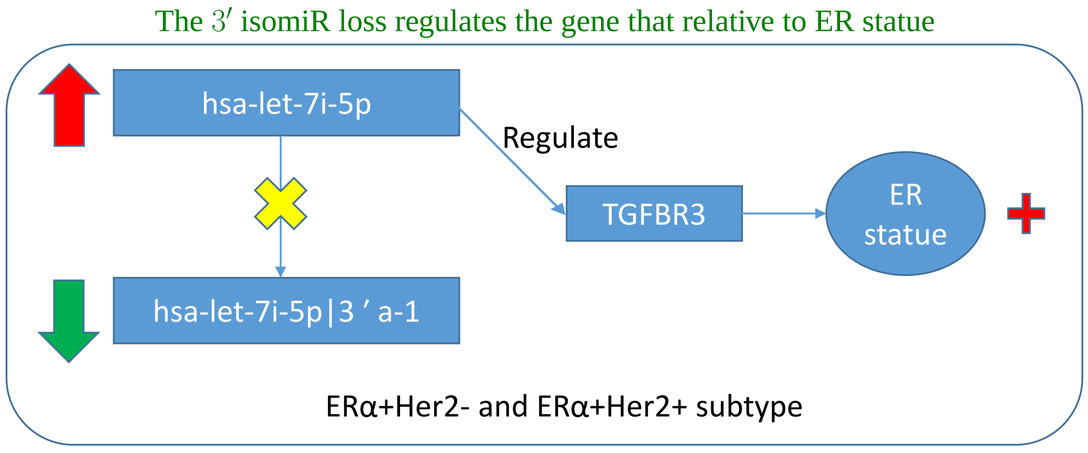
Results

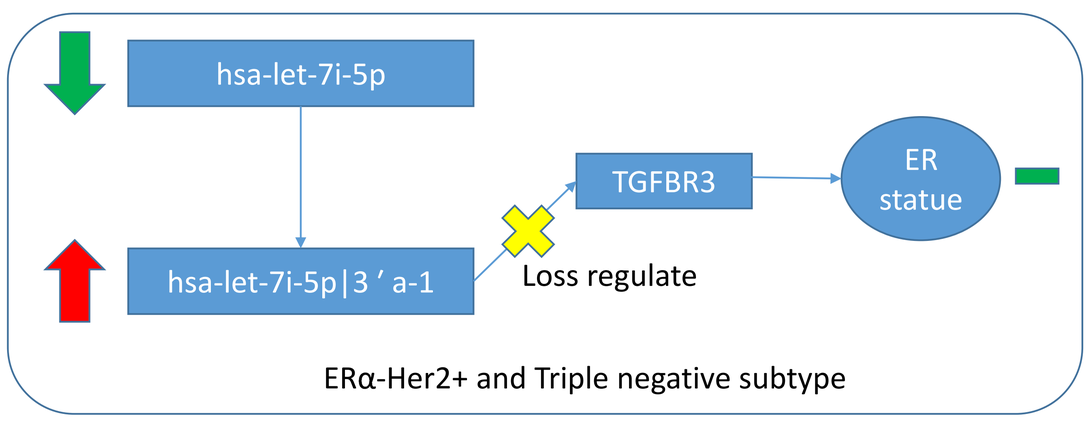
- 20134 different isomiRs were identified in 698 breast cancer patients.
- 435 isomiRs were highly expressed in breast cancer patients.
- 169 isomiRs were $5'$ isomiRs and 266 isomiRs were $3'$ isomiRs. These isomiRs are derived from 169 wild type miRNAs.

Overall view on isomiRs in breast cancer subtypes: (a) $3'$ isomiRs are more abundant than wild type miRNAs. (b) $5'$ isomiRs have comparable expression level to the wild type miRNAs.



The $5'$ isomiR loss regulates the gene that relative to ER statue





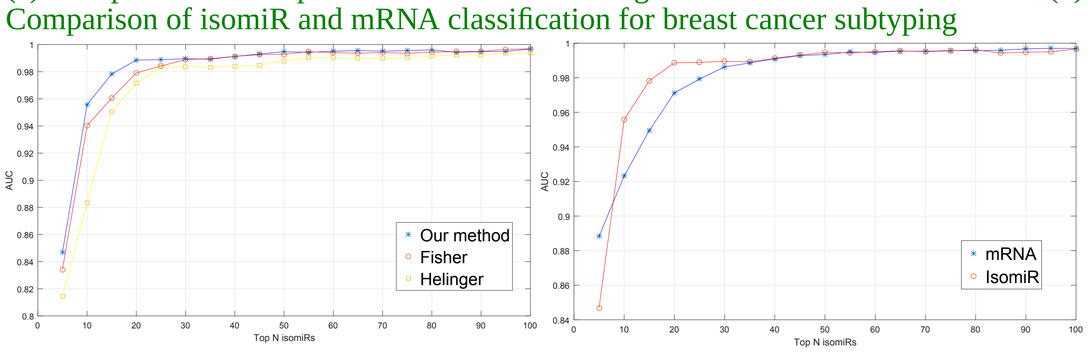
The $5'$ isomiRs may regulate genes that are likely specify breast cancer subtypes

isomiR	Predicted targets	Pathway	Effect
hsa-miR-93-5p|5$'$t-1	SHC4	Estrogen signalling	Estrogen statue
hsa-mir-27a-3p|5$'$t-1	MAPK14	MAPK signaling	Herceptin statue
hsa-miR-92a-1-3p|5$'$t-1	MAPK8	MAPK signaling	Herceptin statue
hsa-mir-106b-3p|5$'$t-1	RAP1B	MAPK signaling	Herceptin statue

Five KEGG pathways which are relevant in breast cancer progression and subtype specification

KEGG name	P-value	Number of gene	Effect
Pathways in cancer	$5.01 * 10^{-11}$	96	Breast cancer outcome
p53 signalling pathway	$1.29 * 10^{-6}$	24	Breast cancer outcome
MAPK signalling pathway	$1.20 * 10^{-5}$	56	Herceptin statue
Insulin signalling pathway	$3.16 * 10^{-3}$	29	Herceptin statue
Estrogen signalling pathway	$1.79 * 10^{-2}$	20	Estrogen statue

(a) Comparison of the presented method to Helinger method and Fisher method. (b) Comparison of isomiR and mRNA classification for breast cancer subtyping



Conclusions

- IsomiR is highly expressed in breast cancer and may regulate the biological process of breast cancer.
- IsomiR can be used as biomarker for the breast cancer subtype classification.
- The presented method is better than traditional methods for discovering isomiR to distinguish different breast cancer subtypes.

References

[1] Aristeidis G Telonis, Phillipe Loher, Yi Jing, Eric Londin, and Isidore Rigoutsos. Beyond the one-locus-one-mirna paradigm: microrna isoforms enable deeper insights into breast cancer heterogeneity. *Nucleic acids research*, 43(19):9158–9175, 2015.

[2] Aristeidis G Telonis, Rogan Magee, Phillipe Loher, Inna Chervoneva, Eric Londin, and Isidore Rigoutsos. Knowledge about the presence or absence of mirna isoforms (isomirs) can successfully discriminate amongst 32 tcga cancer types. *Nucleic acids research*, 45(6):2973–2985, 2017.

Acknowledgements

This work was funded by China Scholarship Council.