Breast Cancer

The most global cancer incidence in women

<table>
<thead>
<tr>
<th>Rank</th>
<th>Cancer</th>
<th>New cases diagnosed in 2018</th>
<th>% of all cancers (excl. non-melanoma skin cancer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Breast</td>
<td>2,088,849</td>
<td>25.4</td>
</tr>
<tr>
<td>2</td>
<td>Colorectal</td>
<td>794,958</td>
<td>9.7</td>
</tr>
<tr>
<td>3</td>
<td>Lung</td>
<td>725,352</td>
<td>8.8</td>
</tr>
<tr>
<td>4</td>
<td>Cervix uteri</td>
<td>569,847</td>
<td>6.9</td>
</tr>
</tbody>
</table>

• Breast cancer causes the greatest number of cancer-related deaths among women. In 2018, it is estimated that 627,000 women died from breast cancer – that is approximately 15% of all cancer deaths among women.

• While breast cancer rates are higher among women in more developed regions, rates are increasing in nearly every region globally.
## Diversity of Breast Cancer: Subtypes

<table>
<thead>
<tr>
<th>Subtype</th>
<th>Percentage at Diagnosis</th>
<th>Receptor Expression</th>
<th>Treatment Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luminal A</td>
<td>40%</td>
<td>Estrogens and progesterone</td>
<td>Chemotherapy</td>
</tr>
<tr>
<td>Luminal B</td>
<td>20%</td>
<td>HER2 Positive Positive</td>
<td>Anti-HER2 therapies</td>
</tr>
<tr>
<td>HER2 Positive Positive</td>
<td>10-15%</td>
<td>Human epidermal growth factor 2</td>
<td>Hormonal therapies</td>
</tr>
<tr>
<td>Triple Negative</td>
<td>15-20%</td>
<td></td>
<td>Novel target therapies</td>
</tr>
</tbody>
</table>

- **Percentage at diagnosis**: The percentage of patients with each subtype at diagnosis.
- **Receptor expression**:
  - Estrogens and progesterone
  - HER2 Positive Positive
  - Human epidermal growth factor 2

- **Treatment strategies**:
  - Chemotherapy
  - Anti-HER2 therapies
  - Hormonal therapies
  - Novel target therapies

Clinical unmet needs in breast cancer

1. Early detection
2. New progression mechanism for targeted therapies (e.g. TNBC)
3. Heterogeneity
4. Drug Resistance
5. Companion diagnosis
6. Racial disparity
Stage & Diversity of Breast Cancer

**TNM Stage**
- **T** = size of primary tumor
- **N** = the extent of spread to nearby lymph nodes
- **M** = presence or absence of distant metastases

**DCIS:** Ductal carcinoma *in situ*

**NORMAL BREAST DUCT**

**Benign**

**Stage 0**

**Stage I**

**Stage II**

**Stage III**

**Stage IV**

**Five year survival rate**
- Stage I: > 95%
- Stage II: About 90%
- Stage III: About 70%
- Stage IV: About 25%
Diversity of Breast Cancer: Racial Disparity

Figure 6a. Trends in Female Breast Cancer Incidence Rates by Race/Ethnicity, 1975-2014, US

Figure 6b. Trends in Female Breast Cancer Death Rates by Race/Ethnicity, 1975-2015, US
Doctors have always recognized that every patient is unique, and doctors have always tried to tailor their treatments as best they can to individuals. You can match a blood transfusion to a blood type — that was an important discovery. What if matching a cancer cure to our genetic code was just as easy, just as standard? What if figuring out the right dose of medicine was as simple as taking our temperature?" - President Obama, January 30, 2015

Liquid Biopsy:
- Less invasive,
- less costly,
- less risky,
- contain more dynamic information than conventional tissue biopsies.
Trend in Cancer Precision Medicine:
Liquid Biopsy: CTC, ctDNA, Exosome

Variety of biomarkers existed in bodily fluids such as blood, saliva, urine, ascites, etc.

Main areas of Liquid Biopsy:
- Circulating tumor cells (CTC)
- Circulating tumor DNA (ctDNA)
- Exosomes
  - mRNA, Protein, miRNA, IncRNA
miRNA Biomarkers in Liquid Biopsy

MicroRNA (miRNA)
1. Non-coding RNA
2. 17-25 nucleotides
3. Post-transcriptional regulation: base-pair with mRNAs and silence those mRNAs
4. Appear to target about 60% of the genes of human

Evaluation of serum microRNA biomarkers for gastric cancer based on blood and tissue pools profiling: the importance of miR-21 and miR-331

Marek Sierzega*1, Marcin Kaczor2, Piotr Kolodziejczyk1, Jan Kulig1, Marek Sanak2 and Piotr Richter1

miR-331 and miR-21 => gastric cancer

Serum MicroRNA profile in patients with colon adenomas or cancer

Yajie Zhang1, Min Li4, Yijiang Ding3, Zhimin Fan1, Jinchun Zhang5, Hongying Zhang6, Bin Jiang3 and Yong Zhu3

8 miRNAs => colon cancer

Circulating microRNAs in breast cancer: novel diagnostic and prognostic biomarkers

Rimi Hamam, Dana Hamam, Khalid A Alsaleh, Moustapha Kassem, Waleed Zaher, Musaad Alfayez, Abdullah Aldahmash & Nehad M Alajez

miR-331 and miR-21 => gastric cancer
Problems to be solved

1. Identify possible biomarkers for early detection of breast cancer from liquid biopsy (e.g. miRNAs).
2. Generate profiling on the various stages of breast cancer. Identify more insightful information regarding the critical factors that progress cancer to the next stage.

Comprehensive data (containing all variation)
+ High-throughput techniques (high sensitivity)
+ Proper data analysis (low sample number issue)
+ Artificial intelligence for optimization (predictable)
Current biomarkers for breast cancer

Breast Cancer: 3 tumor markers
• cancer antigen 15-3 (CA 15-3),
• cancer antigen 27.29 (CA 27.29), and
• carcinoembryonic antigen (CEA)

<table>
<thead>
<tr>
<th>Tumor Marker</th>
<th>N (Total)</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA 15-3</td>
<td>35 (145)</td>
<td>24.1%</td>
</tr>
<tr>
<td>CA 27.29</td>
<td>37 (145)</td>
<td>25.5%</td>
</tr>
<tr>
<td>CEA</td>
<td>27 (145)</td>
<td>18.6 %</td>
</tr>
</tbody>
</table>

## Comprehensive data

Wen-Hong Kuo, MD; PhD

<table>
<thead>
<tr>
<th>Subtypes</th>
<th>Cancer</th>
<th>Stage</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>Luminal A</td>
<td>8</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Luminal B</td>
<td>8</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>HER2 +</td>
<td>8</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Triple Negative</td>
<td>8</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>32</td>
<td>32</td>
<td>48</td>
</tr>
</tbody>
</table>
High-throughput techniques

1. miRNA extraction from serum
2. miRNA analysis with microarray

Takahiro Ochiya, PhD
High-throughput techniques

1. miRNA extraction from serum
2. miRNA analysis with microarray

RNA labeling
RNA 2µL

Fluorescent labeling

DNA chip analysis 3D-Gene®
Hybridization
3cm
miRNA oligo chip (2565 miRNA, miRBase ver. 21)
Wash
7cm

Analysis
Scan
Digitizing

miRNA expression data (Raw data)
DifferenGal expression analysis

ElasGc net (LASSO & Ridge)

Support vector machine (SVM)
Linear discriminant analysis (LDA)
Generalized linear model (GLM)

Modeling
Normalized Data

Feature miRNAs

Imputation (small value)
Calibration with negative controls and internal controls
Differential expression analysis and expression level filtering

Normalization

feature miRNAs

Prediction model

Potential biomarkers & Prediction models

1. Removing
2. Imputation (trimmed mean)

1. Variance stabilizing normalization (VSN)
2. Spike-in VSN
3. CrossNorm

1. Differential expression analysis
2. Elastic net (LASSO & Ridge)
3. Generalized linear model (GLM)

Further Modeling

1. Accuracy Filtering
2. Prediction modeling

Takahiro Ochiya  Chen-An Tsai
Differential Expression Analyses

- **VSN**: 103 probes are selected
- **Spike-in VSN**: 437 probes are selected
- **CrossNorm**: 269 probes are selected

Adjusted P-values < 0.05 (Benjamini–Hochberg procedure)
Significant log2 fold change = 1.5
miRNAs Selected by Elastic Net Regression

Venn Diagram:
- VSN
- Spike-in VSN
- CrossNorm

Numbers:
- 19
- 22
- 32
- 23

Intersection:
- 1

hsa-miR-4783-5p
miRNA Analysis Flow Chart

1. Differential Expression Analysis
2. Elastic net (LASSO & Ridge)

1. Support Vector Machine (SVM)
2. Linear discriminant analysis (LDA)
3. Generalized linear model (GLM)

Evaluation:
Leave one out
Cross validation
# Results of miRNA through Analysis Pipeline

<table>
<thead>
<tr>
<th>Whole miRNA</th>
<th>2565</th>
</tr>
</thead>
<tbody>
<tr>
<td>After removing NA</td>
<td>2462</td>
</tr>
<tr>
<td><strong>Normalization</strong></td>
<td><strong>VSN</strong></td>
</tr>
<tr>
<td>Differential Expression</td>
<td>0</td>
</tr>
<tr>
<td>Elastic Net</td>
<td>103</td>
</tr>
<tr>
<td>Modeling with at most 3 selected miRNA or each single miRNAs</td>
<td>32</td>
</tr>
<tr>
<td>-&gt; Construct 17,423,153 models</td>
<td>31</td>
</tr>
<tr>
<td>-&gt; Evaluate with 10-fold cross validation</td>
<td>103</td>
</tr>
</tbody>
</table>
Modeling with Selected miRNAs
Selecting the miRNAs with prediction ability

sensitivity & specificity > 85%

SVM  LDA  GLM
Overlap and Consistency of Each Modeling Method

Selected miRNAs with Prediction Ability

⇒18 miRNAs (overlap from 4 methods)
PCA and heatmap with Selected 18 miRNAs

NA-omitting data with VSN
PCA plots show selected 18 miRNAs fit-in early detection
Further modeling

After previous prediction modeling, we used the union of selection from each pipeline to build more prediction models. Fisher’s linear discriminant analysis (LDA) was performed with each of these miRNA marker or a combination of at most six miRNA markers.

To evaluate the prediction performance, 10-fold cross validation were applied to each model. We separated the data into 10 groups, built the model with 9 groups and used the residual group as testing cohort to calculate the prediction accuracy, sensitivity and specificity. After repeating the estimation process with different testing group 10 times, the average values of each test result were calculated for model evaluation.

The resulting values of the discriminant functions were used to prepare the diagnostic index.

- Index score ≥ 0: breast cancer
- Index score < 0: non-breast cancer or other clinical conditions
Further Modeling (with 1 miRNA)

Modeling with 1 miRNA (hsa-miR-614)

Threshold: 0.9850
Specificity: 0.8656 ± 0.0964
Sensitivity: 0.8700 ± 0.0881
Accuracy: 0.8683 ± 0.0552
Area Under Curve: 0.9320

Prediction score = 11.3262 - 1.5682 x hsa-miR-614
Further Modeling (with 3 miRNAs)

Prediction score = 5.4049 - 1.7271 x hsa-miR-614 + 0.0937 x hsa-miR-42XX + 1.1171 x hsa-miR-61XX
Further Modeling (with 4 miRNAs)

Prediction score = 9.225 - 0.9554 x hsa-miR-614 + 0.8076 x hsa-miR-42xx + 1.4167 x hsa-miR-61xx - 1.9153 x hsa-miR-66XX

Threshold: 1.4590
Specificity: 0.9533 ± 0.0635
Sensitivity: 0.9747 ± 0.0517
Accuracy: 0.9667 ± 0.0419
Area Under Curve: 0.9871
Further Modeling (with 5 miRNAs)

Prediction score = 8.763 - 0.665 x miR-614 + 0.865 x miR-42xx + 1.413 x miR-61xx - 1.697 x miR-66xx - 0.716 x miR-1xxx

Threshold: 1.5280  
Specificity: 0.9611 ± 0.0533  
Sensitivity: 0.9800 ± 0.0396  
Accuracy: 0.9729 ± 0.0326  
Area Under Curve: 0.9885
Validation

• Patient Serum Collection: Healthy, Benign, pre-Cancer & Cancer

• Workflow:
  1. Isolate RNA
  2. Optimize primer
  3. Reverse transcription
  4. Amplify cDNA
  5. Run qPCR / nanostring/(liquid) chip
  6. Analyze
Conclusions

1. While applying different analysis pipeline would get quiet different outcomes, there are some overlaps, which show the consistency of these methods.

2. Several serum miRNAs are enough to identify the group (cancer or noncancer) of a patient at a high accuracy level. Thus, these selected miRNAs could be viewed as potential biomarkers for implementing early detection of breast cancer.
Perspectives

1. The models seem to be precise enough to fit early detection, more validations are still required to establish robust criteria.
2. The established useful analysis pipeline enables applying for other different expression data derived from other diseases.
3. The mechanisms of these selected miRNAs related are unknown. It is much more meaningful and critical for the understanding of these identified biomarkers. By comprehending the molecular mechanisms underlying these biomarkers, the developing effective treatments and translational research would be promoted.
### Acknowledgements

#### Lab members
- Bao-Hong Lee, PhD
- Yu-Ling Tai, PhD
- Yumi Yasua, PhD
- Emily Lo, PhD
- Shion-Shan Shen, PhD
- Shang-Jer Kuo (PhD student)
- Chia-Yu Yang (Master)
- Ru-Ying Fang (Master)
- Yi-Ting Chen (Master)
- Dan-Jung Chuang (Master)
- Shi-Chen Wang (Master)
- Mitch Lin (Master)
- Angela Wang (RA)
- Jeff Ku (RA)

#### NTU
- King-Jen Chang, MD, PhD
- Wen-Hong Kuo, MD, PhD
- Ming-Yang Wang, MD, PhD
- Chen-Chih Hsu, PhD
- Chen-An Tsai, PhD

#### Cornell
- David C. Lyden, MD, PhD
- Ayuko Hoshino, PhD
- Héctor Peinado, PhD (CNIO, Spain)

#### Japan
- Takahiro Ochiya, PhD (NCC)
- Hidetoshi Tahara, PhD (U Hiroshima)
- Koji Ueda, PhD (Cancer Found.)

#### Others
- Arthur Lander, MD, PhD (UC Irvine)
- Kuo-Kan Liang (AS)
- Ali Mortazabi, PhD (UC Irvine)
- Zhongmin Zhao, PhD (UT Health)