Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer
Motivation

**Observation**: breast-cancer patients at same stage have different outcomes

**Problem**: existing outcome predictors are poor
- lymph nodes
- histological grade

**Goals**:
- identify categories of breast cancer
- predict outcome based on gene expression
- decide therapy accordingly
“Chemotherapy or hormonal therapy reduces risk of distant metastases by approximately one-third; however 70-80% of patients receiving this treatment would have survived without it.”

A main contribution of this method is lower false positives.
## Result Overview

### Table 1 Breast cancer patients eligible for adjuvant systemic therapy

<table>
<thead>
<tr>
<th>Consensus</th>
<th>Total patient group (n = 78)</th>
<th>Metastatic disease at 5 yr (n = 34)</th>
<th>Disease free at 5 yr (n = 44)</th>
</tr>
</thead>
<tbody>
<tr>
<td>St Gallen</td>
<td>64/78 (82%)</td>
<td>33/34 (97%)</td>
<td>31/44 (70%)</td>
</tr>
<tr>
<td>NIH</td>
<td>72/78 (92%)</td>
<td>32/34 (94%)</td>
<td>40/44 (91%)</td>
</tr>
<tr>
<td>Prognosis profile*</td>
<td>43/78 (55%)</td>
<td>31/34 (91%)</td>
<td>12/44 (27%)</td>
</tr>
</tbody>
</table>

The conventional consensus criteria are: tumour $\geq$ 2 cm, ER negative, grade 2–3, patient $<35$ yr (either one of these criteria; St Gallen consensus); tumour $>1$ cm (NIH consensus).

* Number of tumours having a poor prognosis signature using our microarray profile, defined by the optimized sensitivity threshold in the 70-gene classifier (see Fig. 2b).

† Number of tumours with a poor prognosis signature in the group of disease-free patients, when the cross-validated classifier is applied.
## Result Overview

### Table 1: Breast cancer patients eligible for adjuvant systemic therapy

<table>
<thead>
<tr>
<th>Consensus</th>
<th>Total patient group (n = 78)</th>
<th>Metastatic disease at 5 yr (n = 34)</th>
<th>Disease free at 5 yr (n = 44)</th>
</tr>
</thead>
<tbody>
<tr>
<td>St Gallen</td>
<td>64/78 (82%)</td>
<td>33/34 (97%)</td>
<td>31/44 (70%)</td>
</tr>
<tr>
<td>NIH</td>
<td>72/78 (92%)</td>
<td>32/34 (94%)</td>
<td>40/44 (91%)</td>
</tr>
<tr>
<td>Prognosis profile*</td>
<td>43/78 (55%)</td>
<td>31/34 (91%)</td>
<td>12/44 (27%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(18/44 (41%)†)</td>
</tr>
</tbody>
</table>

The conventional consensus criteria are: tumour ≥2 cm, ER negative, grade 2–3, patient <35 yr (either one of these criteria; St Gallen consensus); tumour >1 cm (NIH consensus).

* Number of tumours having a poor prognosis signature using our microarray profile, defined by the optimized sensitivity threshold in the 70-gene classifier (see Fig. 2b).

† Number of tumours with a poor prognosis signature in the group of disease-free patients, when the cross-validated classifier is applied.
Result Overview

needs therapy

<table>
<thead>
<tr>
<th></th>
<th>St Gallen</th>
<th>NIH</th>
<th>Prognosis Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommend Therapy %</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

does not need therapy

<table>
<thead>
<tr>
<th></th>
<th>St Gallen</th>
<th>NIH</th>
<th>Prognosis Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommend Therapy %</td>
<td>25</td>
<td>75</td>
<td>75</td>
</tr>
</tbody>
</table>
Study

1. unsupervised clustering, look for tumor categories

2. supervised learning, find prognosis reporter genes
Study

1. unsupervised clustering, look for tumor categories

2. supervised learning, find prognosis reporter genes
Unsupervised Hierarchical Clustering: Dendogram

http://youtu.be/XJ3194AmH40?t=5m
2 categories
4 categories
co-regulates with ER-α
co-regulates with lymphocytic infiltrate
Study

1. unsupervised clustering, look for tumor categories

2. supervised learning, find prognosis reporter genes
Study

1. unsupervised clustering, look for tumor categories

2. supervised learning, find prognosis reporter genes
Method

Start with 25K genes (some double counting)

From these, identify 5K that are significantly regulated

From these, identify 231 significantly associated with disease outcome

From these, identify 70 as classification features
Supervised Learning on Prognosis Signatures
Training Data (78 tumors)
Test Data (19 tumors)
Supervised Learning on ER and BRCA1 Signatures
Summary

New ideas (for 826):
- new label: clinical outcome
- use of unsupervised learning
- accuracy vs. sensitivity tradeoffs