Gene expression in cancer diagnostics

Aron Eklund

eklund@cbs.dtu.dk

Cancer Systems Biology group
Center for Biological Sequence Analysis
Technical University of Denmark

Introduction to Systems Biology
February 21, 2012
Cancer

The disease caused by an uncontrolled division of abnormal cells in a part of the body

"Hallmarks of cancer"

Hanahan and Weinberg, Cell (2011)
USA lifetime risk of developing cancer (%)

Female

- Breast
- Lung
- Colorectal
- Uterine corpus
- Ovary
- Cervix
- Thyroid
- Pancreas
- Leukemia
- Stomach
- Oral
- Liver
- Non-Hodgkin Lymphoma
- Melanoma
- Kidney
- Bladder
- Multiple myeloma
- Brain
- Hodgkin disease
- Esophagus
- Larynx
- Testicle
- Prostate

Male

USA lifetime risk of developing cancer (%)

Tuesday, February 21, 2012
Breast cancer

Lifetime risk of developing BC:
- 1 in 8 (women)
- 1 in 769 (men)

Lifetime risk of dying from BC:
- 1 in 36 (women)
- 1 in 3333 (men)
Breast cancer treatment overview

**Event**

**Diagnosis**

**Therapeutic options**

- **Neoadjuvant therapy**
- **Surgical resection**
- **Radiation therapy**
- **Adjuvant therapy**
  - Chemotherapy: 6 months
  - Targeted therapy: 1 year
  - Hormone therapy: 5 years

**Chemotherapy**
- adriamycin
- 5-fluorouracil
- cyclophosphamide
- docetaxel
- methotrexate
- epirubicin
- paclitaxel
- cisplatin

**Targeted therapy**
- trastuzumab
- lapatinib

**Hormone therapy**
- tamoxifen
- aromatase inhibitors

**Reurrence? Metastasis?**

**Death?**
Adjuvant combination chemotherapy: benefit vs. side effects

**Benefit**

**Age < 50 years:**
10% absolute increase in recurrence-free survival after 15 years

**Age 50-69 years:**
3% increase

**Short-term effects**
Emesis
Nausea
Stomatitis
Alopecia
Myelosuppression
Thromboembolism
Myalgias
Neuropathy
Fatigue

**Long-term effects**
Premature menopause/infertility
Weight gain
Cardiac dysfunction
Leukemia


Tuesday, February 21, 2012
Motivation: predictive biomarker discovery

• Many **anti-cancer drugs** are available
  - Effective in a only a fraction of patients
  - Undesirable side effects

**Predictive biomarkers:**
Will a given tumor respond to a particular drug?

• Optimize patient outcome
• Reduce suffering/expense/delay
• Identify drug resistance pathways
What could be a biomarker?

**Protein**
- Expression of a certain protein
- Protein modifications

**mRNA**
- Gene expression levels

**DNA**
- SNPs / mutations
- Copy number of genes
- Methylation

**Other**
- Blood / serum
- Cell morphology
- Tissue characteristics

Examples of predictive biomarkers in cancer

**Protein expression** of the estrogen receptor (ER) in breast cancer
- patient will benefit from anti-estrogen therapy (e.g. tamoxifen, aromatase inhibitors)

**Gene amplification** of ERBB2/HER2 in breast cancer:
- patient will benefit from anti-HER2 therapy (trastuzumab, lapatinib)

**Somatic mutation** of the KRAS gene in colorectal cancer
- patient will *not* benefit from anti-EGFR therapy (cetuximab)

**Germline mutation** in BRCA1 in breast cancer
- patient will benefit from cisplatin (?)

The “Philadelphia chromosome” t(9;22)(q34;q11) **karyotype** in leukemia
- patient will benefit from Bcr-Abl tyrosine kinase inhibitors (imatinib)

Various **gene expression** signatures (Mammaprint, Oncotype DX) in ER-positive breast cancer
- patient will benefit from chemotherapy
  (actually a prognostic biomarker)
Why search for gene expression (mRNA) biomarkers?

1. Cell phenotype is largely determined by protein expression
   - Most phenotype-determining molecules are proteins (enzyme, signalling molecules)

2. Proteins are difficult to measure in high-throughput, parallel fashion
   - (but not for lack of trying)

3. mRNA level is an upstream surrogate marker of protein level
   (although not perfectly correlated)

4. Nucleic acids (DNA, RNA) are easy to quantify in parallel

Still: it is possible to discover a biomarker based on gene expression
.. and then implement a clinical test based on protein level
How to measure gene expression (mRNA)?

1. 1970s - Northern blot
   - one gene at a time, not very accurate

2. 1980s - Quantitative reverse transcriptase polymerase chain reaction (RT-PCR)
   - one gene at a time, very accurate

3. mid-1990s - DNA microarray
   - many genes at a time, medium accuracy

4. Late 2000s - RNA-seq
   - all genes, high accuracy?
Effect of Chemotherapy for Patient Groups Defined by the RS

The Kaplan-Meier plots and estimates of the proportion of patients distant recurrence–free for each of the RS risk categories are shown in Figure 2 and Table 1, respectively. The magnitude of the chemotherapy benefit was greater for the high-risk patients \( (RS/31) \) than for the intermediate- \( (RS, 18 \) to \( 30) \) or low-risk patients \( (RS/18) \). There was a large benefit of chemotherapy in the high-risk patients, whereas there was minimal, if any, benefit of chemotherapy in the low-risk patients. The 10-year Kaplan-Meier estimate for freedom from distant recurrence was improved from 60% to 88% by adding chemotherapy to tamoxifen in the high-risk group.

The relative and absolute benefits of chemotherapy for RS risk groups are shown in Figure 3. Although no demonstrable reduction in distant recurrence at 10 years for the predefined low risk category was evident \( (relative \ risk, 1.31; 95% \ CI, 0.46 \ to \ 3.78; increase \ of \ 1.1\% \ in \ absolute \ risk) \), a large reduction in distant recurrence at 10 years was evident for the high-risk category \( (relative \ risk, 0.26; 95% \ CI, 0.13 \ to \ 0.53; decrease \ of \ 27.6\% \ in \ absolute \ risk) \). The benefit from chemotherapy was less clear for patients in the intermediate-risk group \( (relative \ risk, 0.61; 95% \ CI, 0.24 \ to \ 1.59; 1.8\% \ increase \ in \ absolute \ risk) \). Similar trends were observed for freedom from locoregional and/or distant recurrence and overall survival (Figs A1 and A2, online-only appendix).

Similar results were observed when chemotherapy benefit was analyzed in the CMF and MF groups separately (data not shown).

Relationship of the RS and Chemotherapy Treatment Benefit

To test the statistical strength of the relationship between the magnitude of chemotherapy benefit and RS, a formal test of statistical interaction between the RS as a continuous variable and chemotherapy treatment was performed. In a multivariate analysis...
Example of a gene expression *prognostic* biomarker

**Oncotype DX**

RT-PCR measurements of 21 genes

US $4175 / test

Should chemotherapy be given (in addition to the hormone therapy tamoxifen)?
grade reported in the site's pathology reports. There was a modest concordance between RS and patient age. Although RS was associated with PR by ligand-binding and poor histologic grade, a large numbers of cases were discordant. For example, there were cases of tumors with low RSs that had low PR by ligand-binding or poor tumor grade. In addition, there was only modest agreement in assessment of histologic grade among the three pathologists (Tables A2-A4, online-only appendix).

Effect of Chemotherapy for Patient Groups Defined by the RS

The Kaplan-Meier plots and estimates of the proportion of patients distant recurrence-free for each of the RS risk categories are shown in Figure 2 and Table 1, respectively. The magnitude of the chemotherapy benefit was greater for the high-risk patients (RS/H11350 31) than for the intermediate- (RS, 18 to 30) or low-risk patients (RS/H11021 18). There was a large benefit of chemotherapy in the high-risk patients, whereas there was minimal, if any, benefit of chemotherapy in the low-risk patients. The 10-year Kaplan-Meier estimate for freedom from distant recurrence was improved from 60% to 88% by adding chemotherapy to tamoxifen in the high-risk group.

The relative and absolute benefits of chemotherapy for RS risk groups are shown in Figure 3. Although no demonstrable reduction in distant recurrence at 10 years for the predefined low risk category was evident (relative risk, 1.31; 95% CI, 0.46 to 3.78; increase of 1.1% in absolute risk), a large reduction in distant recurrence at 10 years was evident for the high-risk category (relative risk, 0.26; 95% CI, 0.13 to 0.53; decrease of 27.6% in absolute risk). The benefit from chemotherapy was less clear for patients in the intermediate-risk group (relative risk, 0.61; 95% CI, 0.24 to 1.59; 1.8% increase in absolute risk). Similar trends were observed for freedom from locoregional and/or distant recurrence and overall survival (Figs A1 and A2, online-only appendix).

Similar results were observed when chemotherapy benefit was analyzed in the CMF and MF groups separately (data not shown).
Example of a *predictive* biomarker

**Cetuximab** - monoclonal antibody against EGFR (epidermal growth factor receptor)

Approved for colorectal cancer and head and neck cancer

*Why is it effective in some patients and not others?*
A model for cetuximab resistance

A model for cetuximab resistance

<table>
<thead>
<tr>
<th>No drug</th>
<th>Drug with intact pathway</th>
<th>Drug with mutated pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Diagram A" /></td>
<td><img src="image2" alt="Diagram B" /></td>
<td><img src="image3" alt="Diagram C" /></td>
</tr>
</tbody>
</table>

**Diagram A**
- No drug
- EGFR, Ligand, Ras, Raf, Mek, ERK/MAPK
- Transcription factors

**Diagram B**
- Drug with intact pathway
- EGFR, Cetuximab, Ligand, Ras, Raf, Mek, ERK/MAPK
- Transcription factors

**Diagram C**
- Drug with mutated pathway
- EGFR, Cetuximab, Ras, Raf, Mek, ERK/MAPK
- Transcription factors

Khambata-Ford et al. (2007) JCO
K-ras mutation predicts lack of benefit from cetuximab in metastatic colorectal cancer

Patients with mutated KRAS

No benefit

Patients with wild-type KRAS

Improved survival in patients given cetuximab

Karapetis et al. (2008) *NEJM*
Established predictive biomarkers in breast cancer:

1. Estrogen receptor (ER) expression predicts benefit from tamoxifen

2. HER2 expression/amplification predicts benefit from trastuzumab (Herceptin)

What about biomarkers/predictors for other drugs?

How do we discover a predictive biomarker?
Predictive biomarker discovery in a neoadjuvant clinical trial

1. Diagnosis
2. Tumor measurement
3. Core biopsy
4. Drug treatment
5. Tumor measurement
6. Surgery and/or additional treatment as appropriate

Did the tumor *respond* (shrink) or *progress* (grow)?
The gene expression microarray

Labeled target
RNA made from a biological sample

Specific hybridization
between a complementary probe and target
detected by fluorescence

DNA probe
attached to the microarray substrate

Measured intensity ~ gene transcript abundance
... for ~10^4 genes
“expression profile”
Gene expression profiles of breast cancer

One could ask:
Are any of these genes correlated with drug response?
The problem with associative studies

20 - 300 patients
\times
20,000 - 60,000 noisy genes
= 
Many false positives!

Therefore we must reduce our hypothesis space (the number of genes considered)
Identification of candidate biomarkers to predict paclitaxel response

siRNA screen for genes that modulate paclitaxel sensitivity \textit{in vitro}

Derivation of a paclitaxel response metagene

RNAi screen

Functional evaluation

Expression data

Coherent modules

Pro-response

RNAi screen (779 genes)

Mitotic genes (10)
BUB1B, CDAD1, MASTL
CDC2, CSNK1A1L, STK4
TTK, EEF2K, SCYL1

Anti-response

Ceramide RNAi library (50 genes)

Ceramide genes (4)
UGCG, GBA1
COL4A3BP, GBA3

Mitotic genes (6)
BUB1B, CDC2, AURKB
TTK, CDAD1, STK4

Ceramide genes (3)
UGCG, COL4A3BP, GBA3

Mitotic module (4)
BUB1B, CDC2, AURKB
TTK

Ceramide module (2)
UGCG, COL4A3BP

Paclitaxel response metagene

N. Juul et al., Lancet Oncology 11:358 (2010)
Testing the paclitaxel response metagene

Triple-negative breast tumors treated with:

- T-FAC ($n = 27$)
- FAC ($n = 48$)

AUC = 0.79

Taxol (paclitaxel)
5-Fluorouracil
Anthracycline
Cyclophosphamide

**Paclitaxel** response in ER-/PR-/HER2- (triple-negative) tumors

Logistic regression

<table>
<thead>
<tr>
<th></th>
<th>Univariate analysis</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All patients</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oestrogen receptor</td>
<td>0·084; 0·04–0·19; &lt;0·0001*</td>
<td>0·18; 0·07–0·47; 0·00039*</td>
</tr>
<tr>
<td>HER2</td>
<td>3·84; 1·84–8·01; 0·00037*</td>
<td>2·35; 1·01–5·47; 0·048*</td>
</tr>
<tr>
<td>T stage</td>
<td>1·07; 0·55–2·08; 0·84</td>
<td>0·67; 0·29–1·52; 0·34</td>
</tr>
<tr>
<td>Ki67</td>
<td>3·60; 1·79–7·23; 0·00032*</td>
<td>1·17; 0·45–3·07; 0·75</td>
</tr>
<tr>
<td>Grade</td>
<td>6·24; 2·77–14·04; &lt;0·0001*</td>
<td>1·99; 0·73–5·44; 0·18</td>
</tr>
<tr>
<td>Node</td>
<td>2·42; 1·07–5·48; 0·034*</td>
<td>1·98; 0·77–5·06; 0·15</td>
</tr>
<tr>
<td>Metagene</td>
<td>7·28; 3·23–16·41; &lt;0·0001*</td>
<td>1·79; 0·56–5·70; 0·32</td>
</tr>
<tr>
<td><strong>Triple-negative</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T stage</td>
<td>0·79; 0·26–2·44; 0·68</td>
<td>0·39; 0·09–1·60; 0·19</td>
</tr>
<tr>
<td>Ki67</td>
<td>0·79; 0·27–2·29; 0·66</td>
<td>0·10; 0·01–0·80; 0·030*</td>
</tr>
<tr>
<td>Grade</td>
<td>3·46; 0·67–17·83; 0·14</td>
<td>6·59; 0·76–57·36; 0·088</td>
</tr>
<tr>
<td>Node</td>
<td>3·46; 0·67–17·83; 0·14</td>
<td>9·62; 1·02–90·28; 0·048*</td>
</tr>
<tr>
<td>Metagene</td>
<td>4·51; 1·41–14·43; 0·011*</td>
<td>19·92; 2·62–151·57; 0·0039*</td>
</tr>
</tbody>
</table>

Data are odds ratio; 95% CI; p value. All covariates are binary. The binary metagene and binary Ki67 is defined as the upper vs lower 50th percentile. T stage is T3–4 vs T1–2. Grade is 3 vs 1 and 2. Node is N+ versus N0. *Significant values.

**Table 3:** Univariate and multivariate analysis for response to treatment (combined analysis)
Systems biology research

How to better interpret gene expression levels?
Pathology in cancer treatment

From a tumor specimen:
- Cellular/tissue morphology
- Protein expression
- Gene copy number (FISH)
- Karyotype
- Gene sequence
- Gene expression (RT-PCR, microarray, RNA-seq)
- Copy number profile (SNP array)
- Gene CpG methylation
Complex cellularity of tumor biopsies

“other stromal” = fibroblasts, endothelials, histocytes, adipocytes

Matrix factorization of gene expression profiles into contributions from various modules

Patients

Genes

Measured expression = Module 1 + Module 2 + Module 3 + Module 4 + ...
Matrix factorization: top 10 modules in breast cancer

<table>
<thead>
<tr>
<th>Module</th>
<th>Predicts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Estrogen receptor</td>
<td>Tamoxifen, AIs</td>
</tr>
<tr>
<td>2 T cell</td>
<td></td>
</tr>
<tr>
<td>3 Fibroblast / ECM</td>
<td>FEC?</td>
</tr>
<tr>
<td>4 Prolif / mitosis / CIN</td>
<td>Paclitaxel?</td>
</tr>
<tr>
<td>5 B cell</td>
<td></td>
</tr>
<tr>
<td>6 Degradation / 3’ bias</td>
<td></td>
</tr>
<tr>
<td>7 HER2</td>
<td>Trastuzumab</td>
</tr>
<tr>
<td>8 Apocrine</td>
<td></td>
</tr>
<tr>
<td>9 S100</td>
<td></td>
</tr>
<tr>
<td>10 HPGD</td>
<td></td>
</tr>
</tbody>
</table>

~26 in all
Outlook

* Medicine is becoming more personalized / stratified

* High throughput measurements lead to improved understanding of cancer mechanisms: development, progression, metastasis, drug resistance -- and ways to target these mechanisms

* Microarrays will probably be replaced by RNA-seq in the near future, but we are not there quite yet