Microarray analyses: From mouse model to predicting human survival

16 Oct 2009
Aixiang Jiang
Cancer Biostatistics Center
Vanderbilt University School of Medicine

Supported by: USPHS grants CA 69457, T32 CA106183, The VMC CTSA, The Vanderbilt Medical-Scientist Training Program, GI SPORE and the 2008-2009 SUS-Ethicon Scholarship Grant Award
Other team members

- Josh Smith
- Bing Zhang
- Pengcheng Lu
- Yu Shyr
- ......
Change in the US Death Rates* by Cause, 1950 & 2005

* Age-adjusted to 2000 US standard population.
Sources: 1950 Mortality Data - CDC/NCHS, NVSS, Mortality Revised.
The problem

- Colon cancer was estimated to be the 2\textsuperscript{nd} leading cause of cancer-related death in the US for 2008.

- Although a number of groups have identified prognostic gene signatures in colon cancer, few have been based upon the biology of \textit{metastasis}.
Mouse model

- MC-38 cells were transfected with firefly luciferase gene and selected in 0.5mg/mL G418.

- To enrich for invasive MC-38 cells, 7.5 x 10^5 cells were seeded onto 6-well, 8.0 µM pore transwell polycarbonate membrane inserts coated with 2.5 mg/mL matrigel and incubated with serum-free DMEM in the upper chamber and complete DMEM in the bottom well.

- Invading cells were collected after six serial passages through matrigel-coated Boyden chambers.

- The selected invasive cells and the parental luciferase-expressing MC-38 cells were injected into the tail vein and the development of lung metastases was assessed. Development of metastases was determined by bioluminescence imaging.

- The MC-38met cells were then derived by culture of tumor cells from a metastatic lung tumor.
Murine model of metastasis

**In vitro selection (6x)**

MC-38 parental

**In vivo selection**

Tail vein injection (MC-38inv)

Lung nodules

MC-38inv

MC-38met
In vivo monitoring and ex vivo proof of metastases

MC-38 parental

MC-38inv

Days post-injection
1 7 14 21

21 21 21
MC-38met cells are highly metastatic in vivo

Tail-vein injection

Lung

Liver

Splenic injection

MC-38 parental

MC-38met
Development of the recurrence signature - Mouse to Man

MC-38 parental vs. MC-38met
Differential gene expression analysis by microarray

300 Differentially expressed genes with human orthologues
“metastatic gene signature”

PARTIAL TRAINING Set (Vanderbilt Medical Center):
Refinement of 300 homologous genes with high-risk patients
Refinement: Concordance analysis of MC-38met and VMC high-risk patients

“34-gene recurrence signature”

Get model from FULL 55 patient TRAINING Set
(Vanderbilt Medical Center)

TESTING Set (Moffitt Cancer Center): 177 colon cancer patients
Signature from mouse

MC-38 parental vs. MC-38met
Differential gene expression analysis by microarray

300 Differentially expressed genes with human orthologues “metastatic gene signature”

Differentially expressed genes were determined using the limma package in Bioconductor based upon 3 criteria:

1. Fold change > 2;
2. False discovery rate (FDR) based on the moderated t-test followed by Benjamini and Hochberg’s multiple-test adjustment < 0.01;
3. Log odds of differential expression (B-statistic) > 1.
Signature – from Mouse to Man

300 Differentially expressed genes with human orthologues
“metastatic gene signature”

PARTIAL TRAINING Set (Vanderbilt Medical Center):
Refinement of 300 homologous genes with high-risk patients
Refinement: Concordance analysis of MC-38met and VMC high-risk patients

“34-gene recurrence signature”

Directional concordance between MC-38met cells and 19 patients from VMC with high-risk of recurrence or cancer-related death (17-stage IV patients and 2-stage III patients) was determined using a cut-off of $p = 0.08353 \leq 0.10$ (binom.test, at least 13 out of 19 patients have same direction changing as in MC-38 met cells) to refine the metastasis-high-risk signature.
Predicting human survival

“34-gene recurrence signature”

Get model from FULL 55 patient TRAINING Set
(Vanderbilt Medical Center)

TESTING Set (Moffitt Cancer Center): 177 colon cancer patients

Based on full training dataset, run Cox for each Affymetrix probe (34 genes -> 60 Affymetrix probes), get beta and Wald test stat, use these sign of beta and Wald, and expression testing dataset to calculate compound score for each patient in the test dataset, then use compound scores to predict human survival.
Functional genomic clustering analysis (VMC 55)

Stage I
Stage II
Stage III
Stage IV

C6orf64
TMEM14A
TACC2
NQO1
SPRY4
EGR1
VDR
STOX2
HES1
ACYP2
DCTD
AK1
PDLIM5
ACTB
CXR7
MGP
PRTN3
C20orf74
HS3ST5
SLC25A30
MYOT
TEX11
CRABP1
DFNB31
MMP13
CSN3
SPDYA
CIRBP
DENND2A
S100A3
NMNAT3
MUM1L1
SYT17
HPSE

* Recurrence or death

Cluster 1
Cluster 2
### Characterization of the 34-gene recurrence signature

#### Up-regulated Genes

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Fold change</th>
<th>Processes/Networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>CXCR7</td>
<td>3.9</td>
<td>Cancer, survival/growth and chemotaxis</td>
</tr>
<tr>
<td>AK1</td>
<td>2.8</td>
<td>Nucleotide binding</td>
</tr>
<tr>
<td>ACTB</td>
<td>2.8</td>
<td>Cancer, cell morphology and motility, growth, polarization and adhesion</td>
</tr>
<tr>
<td>MGP</td>
<td>2.8</td>
<td>Cell-cell Signaling, branching, migration</td>
</tr>
<tr>
<td>HES1</td>
<td>2.8</td>
<td>Cancer, endocrine function, cell death</td>
</tr>
<tr>
<td>TMEM14A</td>
<td>2.6</td>
<td>Cell proliferation (target of CREB)</td>
</tr>
<tr>
<td>EGR1</td>
<td>2.5</td>
<td>Cancer, endocrine function, cell death</td>
</tr>
<tr>
<td>VDR</td>
<td>2.4</td>
<td>Cancer, endocrine function, cell death</td>
</tr>
<tr>
<td>C6orf64</td>
<td>2.4</td>
<td>Membrane dynamics</td>
</tr>
<tr>
<td>NQO1</td>
<td>2.4</td>
<td>Cancer, cell death</td>
</tr>
<tr>
<td>STOX2</td>
<td>2.3</td>
<td>Putative stem cell marker</td>
</tr>
<tr>
<td>ACPY2</td>
<td>2.2</td>
<td>Mutated in aromatic rice</td>
</tr>
<tr>
<td>SPRY4</td>
<td>2.1</td>
<td>Cancer; cell migration, proliferation, differentiation</td>
</tr>
<tr>
<td>DCTD</td>
<td>2.1</td>
<td>Nucleotide biosynthesis</td>
</tr>
<tr>
<td>TACC2</td>
<td>2.1</td>
<td>Cancer; biogenesis, morphology, proliferation</td>
</tr>
<tr>
<td>PDLIM5</td>
<td>2.0</td>
<td>Cancer, actin binding</td>
</tr>
</tbody>
</table>

#### Down-regulated genes

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Fold change</th>
<th>Processes/Networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRABP1</td>
<td>-21.0</td>
<td>Cancer, Cpg Island Methylation</td>
</tr>
<tr>
<td>MMP13</td>
<td>-17.3</td>
<td>Cell-cell signaling, immune response</td>
</tr>
<tr>
<td>MYOT</td>
<td>-7.1</td>
<td>Cancer, actin filaments and stress fibers</td>
</tr>
<tr>
<td>DFNB31</td>
<td>-5.1</td>
<td>Cell-cell Signaling</td>
</tr>
<tr>
<td>HPSE</td>
<td>-4.7</td>
<td>Cell-cell signaling, immune response</td>
</tr>
<tr>
<td>TEX11</td>
<td>-3.7</td>
<td>Cell cycle/division</td>
</tr>
<tr>
<td>SYT17</td>
<td>-3.1</td>
<td>Membrane protein</td>
</tr>
<tr>
<td>MUM1L1</td>
<td>-2.8</td>
<td>Unknown</td>
</tr>
<tr>
<td>SLC25A30</td>
<td>-2.6</td>
<td>Oxidative stress</td>
</tr>
<tr>
<td>CSN3</td>
<td>-2.5</td>
<td>Cell cycle, membrane dynamics</td>
</tr>
<tr>
<td>NMNAT3</td>
<td>-2.4</td>
<td>Nucleotide biosynthesis</td>
</tr>
<tr>
<td>DENND2A</td>
<td>-2.4</td>
<td>Unknown</td>
</tr>
<tr>
<td>CIRBP</td>
<td>-2.3</td>
<td>Cancer, nucleotide binding</td>
</tr>
<tr>
<td>SPDYA</td>
<td>-2.3</td>
<td>Putative cell cycle</td>
</tr>
<tr>
<td>S100A3</td>
<td>-2.2</td>
<td>Cancer</td>
</tr>
<tr>
<td>PRTN3</td>
<td>-2.1</td>
<td>Cell-cell signaling, immune response</td>
</tr>
<tr>
<td>C20orf74</td>
<td>-2.1</td>
<td>GTPase regulation</td>
</tr>
<tr>
<td>HS3ST5</td>
<td>-2.0</td>
<td>Putative epigenetic regulation</td>
</tr>
<tr>
<td>Study Demographics</td>
<td>VMC-training</td>
<td>MCC-testing</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
<td>55</td>
<td>177</td>
</tr>
<tr>
<td><strong>Mean Age (s.d.)</strong></td>
<td>62.3 (14.1)</td>
<td>65.5 (13.1)</td>
</tr>
<tr>
<td><strong>Sex (%male)</strong></td>
<td>30 (54.5%)</td>
<td>96 (54.2%)</td>
</tr>
<tr>
<td><strong>Stage I</strong></td>
<td>4 (7.3%)</td>
<td>24 (13.6%)</td>
</tr>
<tr>
<td><strong>Stage II</strong></td>
<td>15 (27.3%)</td>
<td>57 (32.2%)</td>
</tr>
<tr>
<td><strong>Stage III</strong></td>
<td>19 (34.5%)</td>
<td>57 (32.2%)</td>
</tr>
<tr>
<td><strong>Stage IV</strong></td>
<td>17 (30.9%)</td>
<td>39 (22%)</td>
</tr>
<tr>
<td><strong>Median Follow-up in Months (Min/Max)</strong></td>
<td>50.2 (0.4 - 111.3)</td>
<td>48.1 (0.92 - 142.6)</td>
</tr>
<tr>
<td><strong>Number of deaths</strong></td>
<td>20 (36.3%)</td>
<td>73 (41.2%)</td>
</tr>
<tr>
<td><strong>Caucasian (%)</strong></td>
<td>50 (90.9%)</td>
<td>151 (85.3%)</td>
</tr>
<tr>
<td><strong>Black (%)</strong></td>
<td>4 (7.3%)</td>
<td>9 (5.1%)</td>
</tr>
<tr>
<td><strong>Other (%)</strong></td>
<td>1 (1.8%)</td>
<td>17 (9.6%)</td>
</tr>
</tbody>
</table>
Overall and disease-specific survival in the Moffitt test set

All Stages: Overall survival

Recurrence score comparison:
Low score (n=88) = 27 deaths
High score (n=89) = 46 deaths

All Stages: Disease-specific survival

Recurrence score comparison:
Low score (n=88) = 22 deaths
High score (n=89) = 33 deaths

n=177
p = 0.003
n=177
p = 0.04
Permutation test

Distribution of 10,000 permutation Wald tests of MCC data with 34-gene poor-prognosis score
Competing Risk:
death as event 1 and cancer as event 2

- HighCompoundScore 1
- LowCompoundScore 1
- HighCompoundScore 2
- LowCompoundScore 2

$P$ of event 1 = $2.8e-06$
$P$ of event 2 = $0.4829$

Months

Probability
The 34-gene score is an independent predictor of recurrence risk

<table>
<thead>
<tr>
<th></th>
<th>Adjusted Hazard Ratio</th>
<th>p-value</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recurrence score</td>
<td>1.016</td>
<td>&lt;0.001</td>
<td>1.008</td>
<td>1.025</td>
</tr>
<tr>
<td>Gender</td>
<td>1.011</td>
<td>0.98</td>
<td>0.481</td>
<td>2.124</td>
</tr>
<tr>
<td>Stage</td>
<td>2.119</td>
<td>0.002</td>
<td>1.312</td>
<td>3.424</td>
</tr>
<tr>
<td>Age</td>
<td>1.001</td>
<td>0.93</td>
<td>0.974</td>
<td>1.029</td>
</tr>
<tr>
<td>Grade</td>
<td>1.446</td>
<td>0.32</td>
<td>0.701</td>
<td>2.985</td>
</tr>
</tbody>
</table>
Relative risk of cancer-related death by percentile score
Summary of Results

• A 34-gene signature was identified using a biological model of metastasis
• The signature was an independent predictor of survival and recurrence in multivariate models
Conclusions

• A biologically-based mouse model identified a stage-independent gene expression signature predictive of poor prognosis in patients with colon cancer.
The current research is accepted by *Gastroenterology*.

*Gastroenterology* is ranked 1st of 55 journals in the Gastroenterology and Hepatology category on the 2008 Journal Citation Reports®, published by Thomson Reuters, and has an Impact Factor of 12.591.
Acknowledgements

**Beauchamp Lab members:**
- Josh Smith
- Natasha Deane
- Nipun Merchant
- Fei Wu
- Jenny Zi
- Tanner Freeman
- Christian Kis
- John Neff
- Nicole Al-Greene

**Biostatistics:**
- Yu Shyr
- Pengcheng Lu

**Bioinformatics:**
- Bing Zhang

**Outside Collaborators:**
- Timothy J. Yeatman, MCC
- Steven Eschrich, MCC
- Kay Washington

**Funding:**
- CA 69457, DK52334, TL1 RR024978 and CA95103
- VMC MSTP-CIT (CTSA)
  - Dr. Brown
  - Dr. Dermody
- SUS Ethicon Scholarship

**R. J. Coffey lab:**
- Ramona Deal
- Jeff Franklin

**Goldenring lab:**
- Joseph Roland
“Few diseases have the power of inspiring fear to the same degree as cancer. . . . . . It is therefore natural that we should strive to throw light upon its nature; but the road to this discovery is both long and difficult”
Professor W. Wernstedt, Dean of the Royal Caroline Institute
- 10 December 1927 prior to the presentation of the 1926 Nobel Prize in Physiology or Medicine to Johannes Fibiger
Last workshop in 2009:

November 20: Fei Ye, PhD

Statistical practice in high-throughput siRNA/shRNA screens to identify genes mediating sensitivity to chemotherapeutic drugs