High-dimensional biomarkers as treatment modifiers in randomized clinical trials

Stefan Michiels, PhD
Head of Methodology
Department of Biostatistics and Epidemiology
Gustave Roussy, Paris-Sud University, Villejuif, France
stefan.michiels@gustaveroussy.fr
I have no financial relationships to disclose.

I will not discuss off label use and/or investigational use in my presentation.
Lessons from early breast cancer: prognostic signatures

Available gene signatures for a price of 400-4175$

- IHC4: 4 genes
- Oncotype Dx: 16 cancer genes
- PAM50 (ROR): 50 genes
- Mammaprint Dx: 70 genes
- Endopredict: 8 cancer genes
- Mapquant Dx (GGI): 97 genes
- ...
...
Instability of gene selection in original Mammaprint training data

Genes included in at least 250 out of 500 (50%) signatures for a training set size of 78 patients

KIAA1750, FGF18, CEGP1, Contig33814_RC, RAMP, EXO1, AL050227, BIRC5, NMU, Contig20217_RC, ANKT, GMPS, Contig38726_RC, Contig55725_RC, PRAME, PEX12, Contig46218_RC, PRC1, HSA250839, CIRBP, PSMD7, HRB, Contig28552_RC, TGFB3

14/70 genes from MammaPrint

10 other genes

Michiels, Koscielny, Hill. Lancet 2005
Association of gene modules with pathological complete response after anthracyclines, beyond clinicopathopathological factors

<table>
<thead>
<tr>
<th>Gene Module</th>
<th>OR</th>
<th>95% CI</th>
<th>P</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGI</td>
<td>1.7</td>
<td>(1.12,2.6)</td>
<td>1.3E-02</td>
<td>3.7E-02</td>
</tr>
<tr>
<td>Gene70</td>
<td>2.02</td>
<td>(1.29,3.2)</td>
<td>2.4E-03</td>
<td>1.3E-02</td>
</tr>
<tr>
<td>CIN70</td>
<td>1.61</td>
<td>(1.08,2.42)</td>
<td>2.1E-02</td>
<td>5.1E-02</td>
</tr>
<tr>
<td>Stroma1</td>
<td>0.73</td>
<td>(0.49,1.06)</td>
<td>1.0E-01</td>
<td>2.1E-01</td>
</tr>
<tr>
<td>Stroma2</td>
<td>0.74</td>
<td>(0.51,1.07)</td>
<td>1.1E-01</td>
<td>2.1E-01</td>
</tr>
<tr>
<td>Immune1</td>
<td>1.92</td>
<td>(1.36,2.73)</td>
<td>2.2E-04</td>
<td>3.7E-03</td>
</tr>
<tr>
<td>Immune2</td>
<td>1.78</td>
<td>(1.25,2.53)</td>
<td>1.3E-03</td>
<td>1.1E-02</td>
</tr>
<tr>
<td>RAS</td>
<td>0.82</td>
<td>(0.57,1.18)</td>
<td>3.0E-01</td>
<td>4.9E-01</td>
</tr>
<tr>
<td>MAPK</td>
<td>0.85</td>
<td>(0.56,1.27)</td>
<td>4.2E-01</td>
<td>6.0E-01</td>
</tr>
<tr>
<td>PTEN</td>
<td>1.75</td>
<td>(1.18,2.62)</td>
<td>5.8E-03</td>
<td>2.5E-02</td>
</tr>
<tr>
<td>AKTmTOR</td>
<td>0.84</td>
<td>(0.59,1.19)</td>
<td>3.2E-01</td>
<td>4.9E-01</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>1.01</td>
<td>(0.67,1.53)</td>
<td>9.5E-01</td>
<td>9.5E-01</td>
</tr>
<tr>
<td>IGF1</td>
<td>0.97</td>
<td>(0.65,1.45)</td>
<td>8.9E-01</td>
<td>9.5E-01</td>
</tr>
<tr>
<td>SRC</td>
<td>1.02</td>
<td>(0.71,1.47)</td>
<td>9.1E-01</td>
<td>9.5E-01</td>
</tr>
<tr>
<td>MYC</td>
<td>1.1</td>
<td>(0.78,1.56)</td>
<td>5.8E-01</td>
<td>7.6E-01</td>
</tr>
<tr>
<td>E2F3</td>
<td>1.6</td>
<td>(1.12,2.3)</td>
<td>1.1E-02</td>
<td>3.7E-02</td>
</tr>
<tr>
<td>BetaCatenin</td>
<td>0.98</td>
<td>(0.68,1.43)</td>
<td>9.4E-01</td>
<td>9.5E-01</td>
</tr>
</tbody>
</table>

OR: odds ratio for a 1-unit change in gene module, adjusted for clinicopathopathological factors
FDR: false discovery rate

Ignatiadis et al JCO 2012
Move to treatment-effect modifiers

Past:
• Development of prognostic signatures by a model

\[
\text{Outcome} \sim \text{biomarker}
\]

• e.g. 99% of published breast cancer signatures

Future:
• Development of “predictive” or treatment-effect modifying signatures

\[
\text{Outcome} \sim \text{biomarker} + \text{treatment} + \text{treatment x biomarker}
\]
Randomize-all design

R: Randomization; Std: standard arm; Exp: experimental arm
Interaction (biomarker-stratified) randomized design

Typically large trials are needed!
AND/OR more sensitive endpoints (such as tumour measurements)

Buyse, Michiels et al, Expert Rev Mol Diag 2011
Most famous subgroup?

ISIS-2: aspirin vs control - effects on vascular death in 17,187 patients with acute myocardial infarction

Peto et al Lancet 1988
Stepwise strategy for high-dimensional data in a randomized clinical trial (RCT)

• **Step 1:** Perform a global interaction test for control of the global Type I error at a prespecified $\alpha$ level (e.g. 5%)

• **Step 2:** Only when global test is significant, develop classifier in a survival model with interaction effects
  – Use 10-fold crossvalidation to estimate treatment effects in composite biomarker score defined subsets (similar to Matsui CCR 2012)
  – Applying survival model on full RCT data: **indication classifier** for future patients (Simon Stat Med 2012)

*Michiels, Rotolo in Matsui, Buyse, Simon 2014*
Step I: Global interaction test by permutation

- Permute the set of biomarkers among the patients, within each treatment arm

- $p$-value = the proportion of $K$ permutations in which the test statistic for global interaction exceeds the test statistic $s$ for the original data

Michiels, Potthoff, George Stat Med 2011
Step II: Develop a composite biomarker classifier through 10-fold crossvalidation

- Divide the RCT in 10 parts

- Part 1
  - Develop biomarker score (training set, 9/10 parts)

- Part 2
  - Evaluate biomarker score (validation set, 1/10 parts)

- Part 3

- Part 10

- Repeat this process 10 times
- Estimate treatment effects according to composite biomarkers scores in the 10 validation sets

Michiels, Rotolo in Matsui, Buyse, Simon 2014
French breast RCTs example

- Tissue-array from two French breast cancer RCTs of adjuvant chemotherapy with long term follow-up (798 pts)
- 11 biomarkers (ER, PR, HER2, EGFR, p53, p27…)
- Disease-free survival, 320 events

Overall Hazard Ratio: 0.78 [0.62–0.97]

<table>
<thead>
<tr>
<th>Years</th>
<th>No chemo</th>
<th>Chemo</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>389</td>
<td>409</td>
</tr>
<tr>
<td>2</td>
<td>335</td>
<td>368</td>
</tr>
<tr>
<td>4</td>
<td>286</td>
<td>321</td>
</tr>
<tr>
<td>6</td>
<td>237</td>
<td>266</td>
</tr>
<tr>
<td>8</td>
<td>195</td>
<td>223</td>
</tr>
<tr>
<td>10</td>
<td>148</td>
<td>161</td>
</tr>
</tbody>
</table>
Step I: Global test, 2000 permutations

- Three different proposed global statistics yield p-values = 0.045, 0.009 and 0.013

Step II: Crossvalidated treatment effects according to composite biomarker score (cut-off at 0.5)

Michiels, Rotolo in Matsui, Buyse, Simon 2014
Prediction of benefit of adjuvant trastuzumab

- Training-validation strategy in NSABP B31 trial, 8-gene signature
- Interaction between trastuzumab and signature: \( p < 0.001 \)

\[ \text{HR} = 1.58 \ (0.67 - 3.69) \quad \text{HR} = 0.60 \ (0.41 - 0.89) \quad \text{HR} = 0.28 \ (0.20 - 0.41) \]

Alternative 1: 
Prospective subset testing

\[ P_{all} < \alpha_{all} \]

- Yes: Claim benefit for all patients
- No:

\[ P_{S+} < \alpha_{S+} \]

- Yes: Claim benefit for S+ patients
- No: No benefit found

S+ : signature-positive patients; all: all patients

*Buyse, Michiels, Curr Op Onc 2013*
Alternative 1: Prospective subset testing

- Simplest approach: split significance level:
  \[ \alpha = \alpha_{\text{all}} + \alpha_{S+} \]
  - the new treatment is compared with the control in the overall population, ignoring the biomarker
  - if \( p_{\text{all}} \leq \alpha_{\text{all}} \) claim effectiveness for all patients
  - if not, the new treatment is compared with the control in biomarker + patients only, and if \( p_{S+} \leq \alpha_{S+} \), claim effectiveness for biomarker + patients only

- There are less conservative, yet properly controlled, ways of adjusting \( \alpha \) for both (correlated) tests

Alternative 2: Two-stage adaptive signature design

1: stage 1 patients
2S+: stage 2 signature-positive patients
all: all patients

- There exists a crossvalidation version (*Freidlin et al CCR 2010*)

*Buyse, Michiels 2013; extension from Freidlin, Simon, CCR 2005*
Pros and contras

• Prospective subset testing: needs prespecified signature

• Gene signature can be developed on a first stage of the trial but statistical power could be small

• Challenge for two-stage adaptive design: possible heterogeneity of treatment effects before and after the adaptation (changes in patient recruitment or other temporal trends)

• Crossvalidation scheme: need for independent validation trial when all data is repeatedly used to develop the signature?
Conclusion

- Move on from *prognostic* gene signatures (trying to be predictive of treatment benefit) to gene signatures **developed on RCT data as treatment modifier**

- Global interaction test at significance level \( a \) by permuting a statistic among the patients within the treatment groups

- **Continuous** gene signature development in RCT by crossvalidation approach

- Strong **challenges in controlling confounding**: handling of specimens, measurement error, tumour heterogeneity, biopsy vs primary vs metastatic specimen
Main references