Prognosis within different breast cancer subtypes using functional activity of signaling pathways

Henk van Ooijen Márcia A. Inda Ralf Hoffmann Paul van de Wiel Anja van de Stolpe Wim Verhaegh

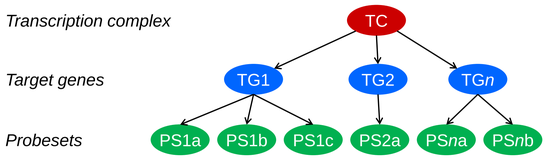
Philips Research, Eindhoven, The Netherlands, contact: wim.verhaegh@philips.com

Summary

- **We developed computational models to assess *functional* activity of the ER, PI3K, AR, Wnt, HH and TGFβ pathways in individual samples, using mRNA expression data.**
- **On 1294 breast cancer samples, 749 (58%) had at least one of the six pathways active, and 167 (13%) at least two.**
- **Functional pathway activity is clearly different across breast cancer subtypes, and adds biological insights.**
- **Functional activity of the modeled pathways is associated with prognosis; combining them in a multi-pathway score (MPS) separates poor from good prognosis cases, complementary to proliferation profiles.**
- **MPS can also predict prognosis within subtypes, showing HER2 cases with relatively poor prognosis and basal cases with relatively good prognosis.**
- **Pathway activities can be used to explain *why* a patient has a poor prognosis, and give suggestions for (targeted) treatment.**

Material & method

We have modeled the transcriptional programs of the ER, PI3K, AR, Wnt, HH and TGFβ pathways, to infer functional pathway activity from mRNA levels of their direct target genes, measured on Affymetrix HG-U133Plus2.0 arrays (fRMA preprocessed). Details of the approach are described in [1]. We modeled the pathways in a probabilistic manner, using a Bayesian network, with three types of nodes: a transcription complex, target genes and probesets. Each model describes (i) how the expression of the target genes depends on the activation of the respective transcription complex, and (ii) how probeset intensities depend in turn on the expression of the respective target genes.



The models can be used to estimate pathway activity in an individual test sample by entering its Affymetrix probeset measurements, and inferring backwards in the model what the probability is that the transcription complex must have been present.

Datasets. Pathway activities were determined on 1294 mixed breast cancer samples from public data sets GSE6532, GSE9195, GSE20685, GSE21653 and E-MTAB-365. Data on relapse-free survival was available for 1169 patients.

Subtypes. Intrinsic breast cancer subtypes were available for GSE21653. Based on those, centroid profiles were calculated for each of the five subtypes using the 50 intrinsic genes from the PAM50 profile [2]. These centroids were used to determine the subtypes of the breast cancer samples from the other data sets.

21-gene recurrence score. A research implementation was made of the 21-gene recurrence score [3] based on microarray data.

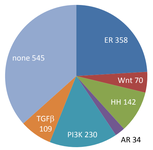
Resulting pathway activities

We assessed the probability of activity of each pathway in each sample. On the 1294 samples, 749 (58%) had at least one pathway active, which is defined as having an inferred probability above 0.5. If we lower the threshold to 0.2 (called marginally active), we get a number of 1026 samples (79%). Furthermore, we often see combinations of activity.

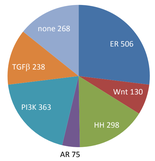
# pathways	active in # patients (%)	marginally active in # patients (%)
0	545 (42%)	268 (21%)
≥ 1	749 (58%)	1026 (79%)
≥ 2	167 (13%)	468 (36%)
≥ 3	26 (2%)	102 (8%)
≥ 4	1 (<1%)	14 (1%)

The ER and PI3K pathways are active most often, as shown below, but the developmental pathways HH, TGFβ and Wnt are also active in a fair share of samples.

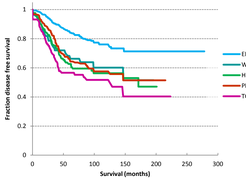
Active pathways (P > 0.5) in samples



Marginally active pathways (P > 0.2) in samples



Prognosis. Pathway activity is associated with relapse-free survival, assessed on the 1169 samples with follow-up info, even though the models were not trained for this purpose. Only AR was not significantly associated.

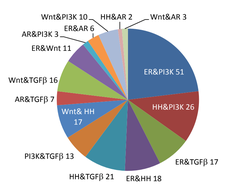


Risk score	Cox coefficient	se	HR	HR 95% CI		p
P_{ER}	-0.87	0.18	0.42	0.29	0.60	9.78E-07
P_{Wnt}	0.38	0.26	1.46	0.88	2.40	0.071
P_{HH}	0.90	0.20	2.45	1.67	3.61	2.72E-06
P_{PI3K}	0.70	0.17	2.02	1.46	2.80	9.97E-06
P_{TGFb}	1.27	0.21	3.56	2.37	5.33	4.01E-10

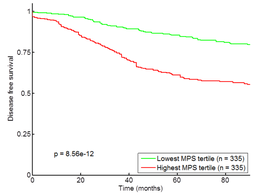
Combinations of pathway activity

In 167 (13%) of the 1294 samples, at least two pathways turned out to be active. Combinations of two pathways are observed as shown below. For prognosis assessment, we hence combine the activities of all pathways for each sample. We do so by weighting them with their univariate Cox coefficients determined on 164 samples from GSE6532 and GSE9195. Testing the resulting multi-pathway score (MPS) on the remaining 1005 samples gives the below Kaplan-Meier curve for the upper and lower tertiles.

Combinations of active pathways in samples

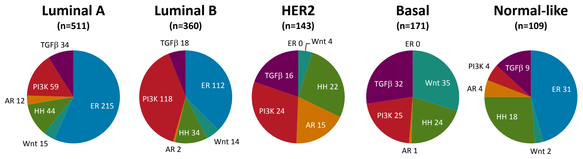


Prognosis with pathway combinations



Activity in breast cancer subtypes

Distribution of pathway activity is clearly different across breast cancer subtypes, as shown below.



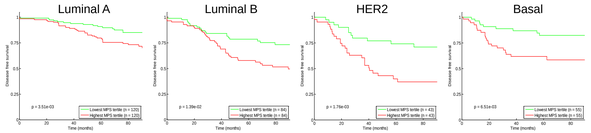
- We see more ER activity in luminal A compared to luminal B, matching the observation that ER activity is associated with better prognosis.
- In luminal B, we see many samples with an active PI3K pathway, also matching the poorer prognosis in this group.
- In the HER2 and basal group, we see no ER activity, as expected, but we do observe more activity of the Wnt, HH and TGFβ pathways.
- The largest fraction in the HER2 group has an active PI3K pathway, as expected.

Although pathway activity is in line with expectations across subtypes, there is still quite a mixture within each subtype. As such, functional pathway activity adds biological insights into a patient's tumor.

Prognosis within subtypes

Our multi-pathway score can also be used to assess prognosis within each breast cancer subtype.

- Although within the luminals, the B type generally have a worse prognosis than the A type, MPS can distinguish even further within the A and B types.
- Remarkably, there is a group of HER2 patients with a very poor prognosis (red line; highest MPS tertile), and a group of basal patients with a fairly good prognosis (green line; lowest MPS tertile).



Combining prognostic profiles

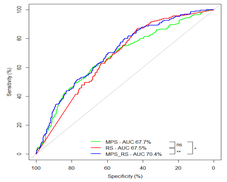
A multivariate analysis of our multi-pathway score (MPS) with the 21-gene recurrence score (RS), shows that the two complement each other when assessed on the 1005 test samples (note: multiple subtypes).

Risk score	Cox coefficient	se	HR	HR 95% CI		p
MPS	1.86	0.36	6.40	3.19	13.0	1.49E-07
RS	0.65	0.17	1.92	1.37	2.69	7.25E-05

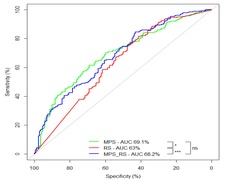
If we plot the ROC curves for predicting 5-year disease recurrence, we see that RS better identifies low risk cases, while MPS better identifies high risk cases, and combining the two gives the best of both profiles. For 10-year recurrence, MPS and the combination show a higher AUC.

In addition to assessing prognosis, MPS can be used to explain *why* a patient is at high risk, and hence may give valuable suggestions for (targeted) treatment.

Disease recurrence < 5 years



Disease recurrence < 10 years



[1] W. Verhaegh et al. Selection of personalized patient therapy through the use of knowledge-based computational models that identify tumor-driving signal transduction pathways. Cancer Res 2014;74(11):2936-45.

[2] J. Parker et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27(8):1160-7.

[3] S. Paik et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol 2006;24(23):3726-34.

get free full text at Cancer Research

This work is supported by NanoNextNL, a micro and nanotechnology consortium of the Government of the Netherlands and 130 partners.