Identifying tumor driving signaling pathways using computational pathway models

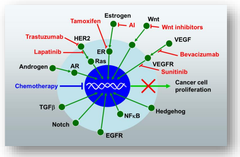
Wim Verhaegh[1] Henk van Ooijen[1] Márcia A. Inda[1] Kalyan Dulla[1] Pantelis Hatzis[2] Rogier Versteeg[3] John Martens[4] John Foekens[4] Hans Clevers[2] Anja van de Stolpe[1]

[1] Philips Research, Eindhoven, The Netherlands, contact: wim.verhaegh@philips.com [2] Hubrecht Institute, Utrecht, The Netherlands [3] AMC, Amsterdam, The Netherlands [4] Erasmus MC, Rotterdam, The Netherlands

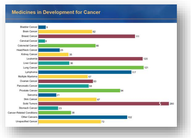
Introduction

A high number of targeted cancer drugs is in development, targeting pathways that drive tumor growth. To select the right targeted cancer drugs for *an individual patient*, one hence needs to know:

- which pathway drives tumor growth in this patient?
- what (epi)genetic defect causes its deregulation?



Targeting tumor driving pathways



PhRMA report 2012: working on 981 medicines for cancer

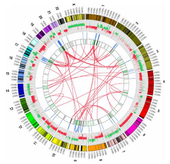
Comprehensive pathway assessment

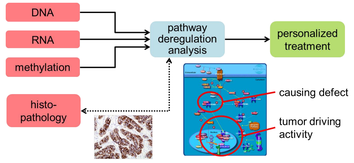
Current tests for therapy selection are not adequate, e.g., breast cancer treatment of ER+ and Her2+ cancers show 25-50% response rates. This can be improved by doing a more comprehensive assessment of the status of oncogenic pathways, measuring

- receptor proteins
- mutations and protein modifications in the signaling cascade
- transcription complex activity
- mRNA expression of target genes.

Interpretation tools

A lot of molecular data is available, e.g. from sequencing, but tools are needed to translate that into meaningful clinical information as to which targeted cancer drug to administer to a patient. To this end, we are developing computational models of the ~10 most relevant oncogenic pathways.

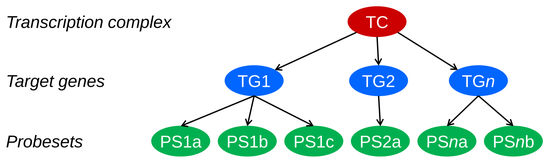




Complex molecular data is mapped onto pathway models to derive useful information

Pathway modeling framework

Most abundantly available data sets concern mRNA microarray measurements, so our initial pathway models cover only the transcriptional program. We model this in a probabilistic manner, by means of a Bayesian network, with three types of nodes: a transcription complex, target genes and probesets. The model describes (i) how the expression of the target genes depends on the activation of the transcription complex, and (ii) how probeset intensities depend in turn on the expression of the respective target genes.



Next, the model can be used to estimate pathway activity by entering the probeset measurements (Affymetrix HG-U133Plus2.0 arrays, fRMA preprocessed), and inferring backwards in the model what the probability must have been for the transcription complex to be present.

Target gene selection

For optimal performance, the models should contain (only) direct target genes of the respective pathways. We selected them based on multiple sources of evidence:

- motif analysis of promotor regions
- transcription factor binding experiments (CHiP-chip/seq/pcr)
- differential expression analysis (activated vs. non-activated pathway)
- input from expert groups if available.

We currently have built four pathway models, for Wnt, ER, AR and Hedgehog, with about 30 target genes each. This number gives specific results on the one hand, and a robust model on the other hand.

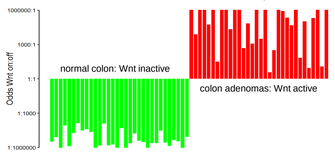
Fitting model parameters

The probabilistic relations in the Bayesian models need to be made quantitative, to allow for quantitative probabilistic reasoning. To improve generalization behavior across tissue types, we have carefully hand-picked the parameters for the upper part of the models. The parameters related to the probeset nodes have been fitted on experimental data:

- cell line experiments with defined active and inactive pathway settings, or
- patient samples with known pathway activity status.

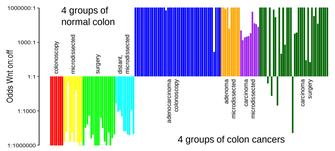
Results - Wnt

Wnt model 1 fitted on Wnt knock-down experiments in colon cell line LS174T gives perfect test results on normal colon and colon adenoma samples:



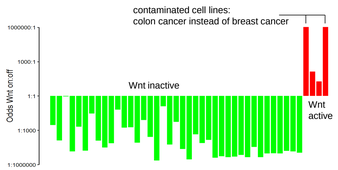
Test set GSE8671 from J. Sabates-Bellver et al., Mol Cancer Res 2007; 5(12):1263-75.

Wnt model 2 fitted on normal colon and colon adenoma samples gives strong separation in colon samples:



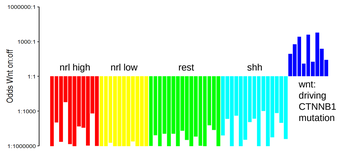
Test set GSE20916 from M. Skrzypczak et al., PLoS ONE 2010; 5(10):e13091.

Wnt model 2 also gives perfect results on breast cancer cell lines, confirmed by a TOPflash reporter contruct:



Test set GSE10890, uploaded by Alex T. Adai, Genentech.
TOPflash reports from A. Hollestelle et al., Breast Cancer Res Treat 2010; 121(1):53-64.

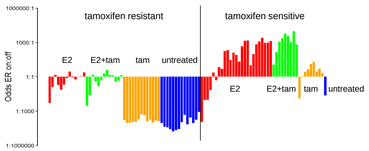
And Wnt model 2 perfectly indentifies medulloblastoma samples with driving β-catenin mutations:



Test set GSE10327 from M. Kool et al., PLoS ONE 2008; 3(8):e3088.

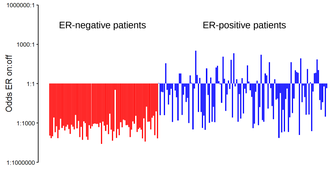
Results - ER

An ER model was fitted on estrogen-deprived and -stimulated MCF7 breast cancer cell lines. It shows a clear difference in activity in tamoxifen-resistant and -sensitive cell lines:



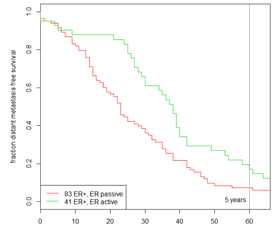
Test samples from GSE21618 from M. Oyama et al., J. Biol Chem 2011; 286(1):818-29. Treatment is between 0 and 48 hours.

ER activity is only predicted in a fraction of ER-positive patient samples:



Test set GSE12276 from P.D. Bos et al., Nature 2009; 459(7249):1005-9.

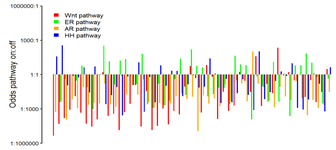
Furthermore, within ER-positive breast cancers, ER activity is associated to better prognosis:



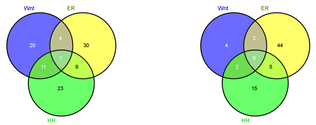
124 ER-positive samples from test set GSE12276 from P.D. Bos et al., Nature 2009; 459(7249):1005-9.
Note: all patients in this study got metastasis, and not all patients have undergone the same treatment regime.

Multiple pathways

By combining multiple pathway scores, for each sample one can assess which pathways are active, in order to understand the biology underlying tumor growth for each patient.

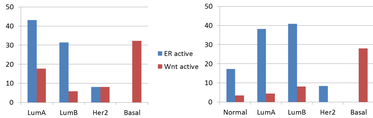


Pathway activities of Wnt, ER, AR and HH for each luminal A sample in test set GSE12276 from P.D. Bos et al., Nature 2009; 459(7249):1005-9.



Combinations of pathway activities in test set GSE12276 (left) from P.D. Bos et al., Nature 2009; 459(7249):1005-9, and combined test sets GSE6532/GSE9195 (right) from S. Loi et al., J Clin Oncology 2007; 25(10):1239-46.

Furthermore, distributions of pathway activity are quite different across breast cancer subtypes:



Percentages of samples with active ER and Wnt pathway across different breast cancer subtypes in test set GSE12276 (left) from P.D. Bos et al., Nature 2009; 459(7249):1005-9, and GSE21653 (right) from R. Sabatier et al., Breast Cancer Res Treat 2011; 126(2):407-20.

Conclusion

- Our initial Wnt model seems to correctly identify an active Wnt pathway across tissue types.
- Our initial ER model shows that ER expression is necessary for ER pathway activity, but not sufficient.
- Within ER+ samples, ER active ones have better prognosis.
- Combining multiple pathway activities sheds light on the biology underlying tumor growth.
- Clinical utility is to be demonstrated next.

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