

Microarray-based expression signatures: potential application for individualized cancer treatment

Citation for published version (APA):

Starmans, M. H. W. (2011). *Microarray-based expression signatures: potential application for individualized cancer treatment*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20111125ms>

Document status and date:

Published: 01/01/2011

DOI:

[10.26481/dis.20111125ms](https://doi.org/10.26481/dis.20111125ms)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



Microarray-based expression signatures:
potential application for individualized cancer treatment

© Copyright Maud Starmans, Maastricht 2011

ISBN 978 94 6159 092 3

Production: Datawyse | Universitaire Pers Maastricht

Microarray-based expression signatures:
potential application for individualized cancer treatment

PROEFSCHRIFT

Ter verkrijging van de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus Prof. mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op vrijdag 25 november 2011 om 14.00 uur

door

Maud Helena Wilhelmina Starmans
geboren te Heerlen op 18 juli 1982



Promotores:

Prof. dr. Ph. Lambin

Prof. dr. B.G. Wouters

Copromotor:

Dr. P.C. Boutros (Ontario Institute for Cancer Research, Toronto, Canada)

Beoordelingscommissie:

Prof. dr. F.C.S. Ramaekers (Voorzitter)

Prof. dr. A.C. Begg (NKI, Amsterdam)

Dr. C.T. Evelo

Prof. dr. M.F. von Meyenfeldt

Dr. Q. Morris (University of Toronto, Toronto, Canada)

Contents

| | | |
|-----------|--|-----|
| Chapter 1 | General introduction | 7 |
| Chapter 2 | Robust prognostic value of a knowledge-based proliferation signature across large patient microarray studies spanning different cancer types | 21 |
| Chapter 3 | The use of a comprehensive tumour xenograft dataset to validate gene signatures relevant for radiation response | 45 |
| Chapter 4 | Validation of a PCR-based proliferation signature <i>in vitro</i> , <i>ex vivo</i> and in patient studies | 59 |
| Chapter 5 | A simple but highly effective approach to evaluate the prognostic performance of gene expression signatures | 81 |
| Chapter 6 | Re: Gene expression-based prognostic signatures in lung cancer: ready for clinical use? | 105 |
| Chapter 7 | Validation of a microarray-based prognostic marker for non-small cell lung cancer: sensitivity to data pre-processing | 109 |
| Chapter 8 | Discussion and future perspectives | 125 |
| Chapter 9 | Summary | 143 |
| | Nederlandse samenvatting | 147 |
| | Acknowledgements/ Dankwoord | 153 |
| | Curriculum Vitae | 159 |
| | List of publications | 163 |

Chapter 1

General introduction

Cancer: a heterogeneous disease

In 2007 an estimated 7.9 million people across the world died from cancer, making it the leading cause of disease-related death worldwide¹ (www.who.int/cancer). The highest incidence of new cancers can be ascribed to breast (women), prostate (men), colon and colorectal, lung and stomach cancers. It is estimated that the global cancer burden worldwide will continue to rise, with a projected 27 million new cases and 17.5 million deaths in 2050².

In the Netherlands almost 90000 new cancer cases were reported in 2008, an increase of 30% compared with 1998 (www.ikcnet.nl). The number of reported deaths also increased and now accounts for 30% (40000 in 2008) of all mortality. Highest numbers of mortality were observed for carcinomas of the lung, colon, breast and prostate^{3,4}.

The standard treatment options available for cancer include surgery, chemotherapy and radiotherapy. Additional treatments can be given for certain cancer types like hormone therapy for breast cancer or immunotherapy for melanoma. Although advances in cancer research have improved prognosis by optimizing administration methods, combining standard therapeutic modalities and developing new drugs, a large fraction of patients do not benefit. Even patients with the same type of cancer have highly variable outcome and survival.

This large variance can be attributed to the fact that cancer is a very heterogeneous disease, it comprises more than 100 distinct types that share a series of common phenotype characteristics that distinguish them from normal cells⁵. Although a series of processes affected during tumorigenesis, like proliferative capacity and growth factor dependency, are similar across patients, the way these characteristics are acquired can be very different⁵. Further, during tumor progression cells can acquire additional properties, for instance increased motility that can initiate invasion and metastasis formation, contributing to tumor aggressiveness⁶. A large number of genes and pathways are involved in these processes and therefore multiple gene mutations are needed for cancer to arise⁷. Recent genome sequencing studies have shown that each tumor has a unique mutational profile, with only a few mutations common to many cancers⁸⁻¹⁰.

Another important aspect of tumor biology that might explain differences in treatment response is the tumor micro-environment⁶. Changes in the physiology of cancer cells lead to fluctuations in nutrient supply, pH, interstitial pressure and

tissue oxygenation that are uncommon for normal tissues¹¹⁻¹⁵. Some of these unique features have previously been linked to treatment efficacy^{11,14-16}. The chaotic and immature organization of tumor blood vessels impairs tumor oxygenation resulting in temporal and regional oxygen fluctuations. Certain tumor areas will be exposed to (severely) reduced oxygen levels, called hypoxia, which is associated with decreased treatment efficacy and enhanced aggressiveness^{16,17}.

Different combinations of mutations and tumor micro-environment give rise to the diverse behavior seen even amongst similar cancer types which is reflected in the highly variable treatment outcome and survival. This elucidates the need for more individualized therapy that takes into account specific genetic and micro-environment tumor properties rather than a one-size fits all strategy.

This leads directly to two pressing challenges in cancer biology. First, how should we select the most appropriate therapy for a specific patient? If molecular markers can be found that predict treatment outcome, these can be used to determine which currently available therapies will be most appropriate. Second how can new targets to create specific drugs for be identified? For example, if a single pathway is affected in poor outcome patients in a variety of different ways, then a drug targeting that pathway as a whole could be helpful. In the last decade, gene expression microarray technology has been applied to address these challenges.

High-throughput Gene Expression Profiling

Briefly, an expression microarray consists of a large panel of probes, representing thousands of genes that are immobilized to a solid matter. Labeled target (RNA, cRNA or cDNA) is subsequently hybridized to these probes and expression levels of each gene present on the microarray can be calculated. A snapshot of the whole transcriptome can thus be assessed in a single experiment.

The ability to analyze the expression of thousands of genes created the opportunity to identify dysregulated pathways and find prognostic and predictive makers. Numerous prognostic gene expression-based markers, or so-called signatures, have been reported since the introduction of this high-throughput technology¹⁸⁻³³.

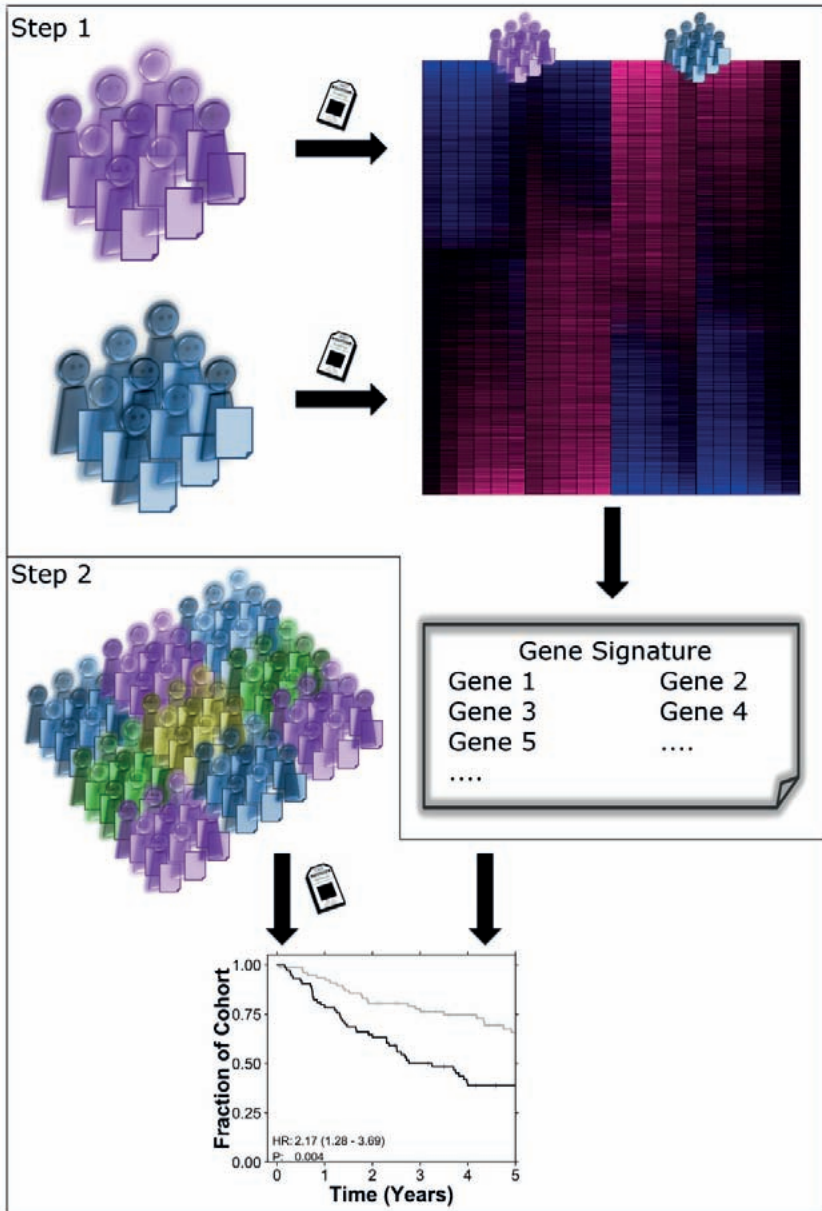


Figure 1.1

Schematic overview of patient data-derived gene-expression signature development. In step 1 gene-expression profiling is performed on distinct (predefined) patient groups to extract genes differentially expressed between the patient groups. These genes are then used as signature and tested in independent patient data for their prognostic and/or predictive power (step 2).

Patient data-based signatures

The most common approach has been to employ gene expression profiling on (large) series of tumor samples. Algorithms combining expression data with clinical parameters and/or survival data are then applied to create gene sets that can identify biologically and prognostically distinct tumor subgroups (Figure 1.1). In breast and non-small cell lung cancer a whole series of gene expression-based markers have been developed since the introduction of this technique^{18,21,23,27,28,32,33}.

The majority of markers are created by comparing gene expression in two patient categories; usually patients who relapse versus patients who do not. In breast cancer the number of created signatures is most extensive.

Using this supervised top-down approach the 70-gene signature was developed from a retrospective series of 78 node-negative breast cancer patients who had received no systemic adjuvant therapy²¹. This marker was validated in a breast tumor series that included both node-positive and node-negative patients³⁴. However since a large subgroup of these patients had received adjuvant systemic treatment and most of the patients in the original study were included in this second series, the results could have been biased. An independent study including more than 300 patients, who did not receive adjuvant systemic therapy, confirmed amongst others the prognostic power of the signature³⁵⁻³⁷. A customized microarray chip (MammaPrint, Agendia) to evaluate the 70-gene signature was recently cleared by the Food and Drug Administration (FDA) to aid in the prognostication of young early-stage node-negative breast cancer patients. The MINDACT trial, a large stage III clinical trial, is ongoing at the moment to confirm the classifier's prognostic power and address its possible predictive value to select patients for systemic adjuvant therapy^{38,39}.

The 21-gene assay (OncoType DX, Genomic Health Inc.) is another gene expression-based marker currently tested in a large stage III clinical trial; the TAILORx trial⁴⁰. This quantitative PCR-based assay was developed for ER-positive breast cancers and can be applied to formalin-fixed paraffin-embedded (FFPE) material²³. First, 250 candidate genes were selected from published literature, which were subsequently profiled with PCR in 3 independent clinical studies comprising more than 400 patients. A 21-gene panel was selected based on its correlation with the likelihood of distant recurrence and was used to develop an algorithm that computes a

recurrence score²³. The 21-gene assay was further validated in multiple studies and a potential predictive capacity was demonstrated⁴¹⁻⁴³.

Perou *et al*¹⁸ applied another approach to identify distinct groups of breast cancers. A series of breast tissues, both normal and malignant, was profiled and diversity in gene expression patterns was related to phenotypic variation in breast tumors. This led to the definition of 'the intrinsic gene list'; a subset of genes that groups breast tumors into distinct molecular subtypes. In a larger follow-up study the subtype classification was extended and differences in survival properties of the different subtypes were revealed⁴⁴. In several consecutive studies the subtypes were further refined and their prognostic implications confirmed⁴⁵⁻⁴⁷.

Biology-based signatures

On the other hand gene expression signatures can be identified based on gene sets involved in a certain biological process or implicated in specific environmental conditions. Usually *in vitro* experiments are performed to create these biology-based gene expression markers. Application of these markers to patient data could potentially be used to identify different prognostic groups and assess tumor sensitivity to a specific drug and predict therapy resistance (Figure 1.2).

Several of these biology-based signatures show prognostic power across different cancer types^{22,26,31,48-52}. One of the earliest examples is the Wound signature²². Similarities between wound healing and tumor microenvironment^{53,54} allowed researchers to the question whether identification of an activated wound response program in cancer could have prognostic implications. An *in vitro* cell culture model for wound healing⁵⁵ was combined with gene expression microarray profiling to create a marker to classify wound response. In short, fibroblasts cultured from different anatomical sites where either cultured in normal media (10% fetal bovine serum (FBS)) or in low-serum media, containing only 0.1% FBS. Genes were selected based on their differential expression between the two conditions. To improve the specificity of their wound response signature, genes directly related to cell proliferation were eliminated from the gene list, leaving the 512 Core Serum Response (CSR) genes. The CSR genes, referred to as the Wound signature, identified subgroups of cancer patients with an activated wound response that correlated with worse survival in breast, lung and gastric cancer²². This prognostic power was validated in multiple studies^{56,57}.

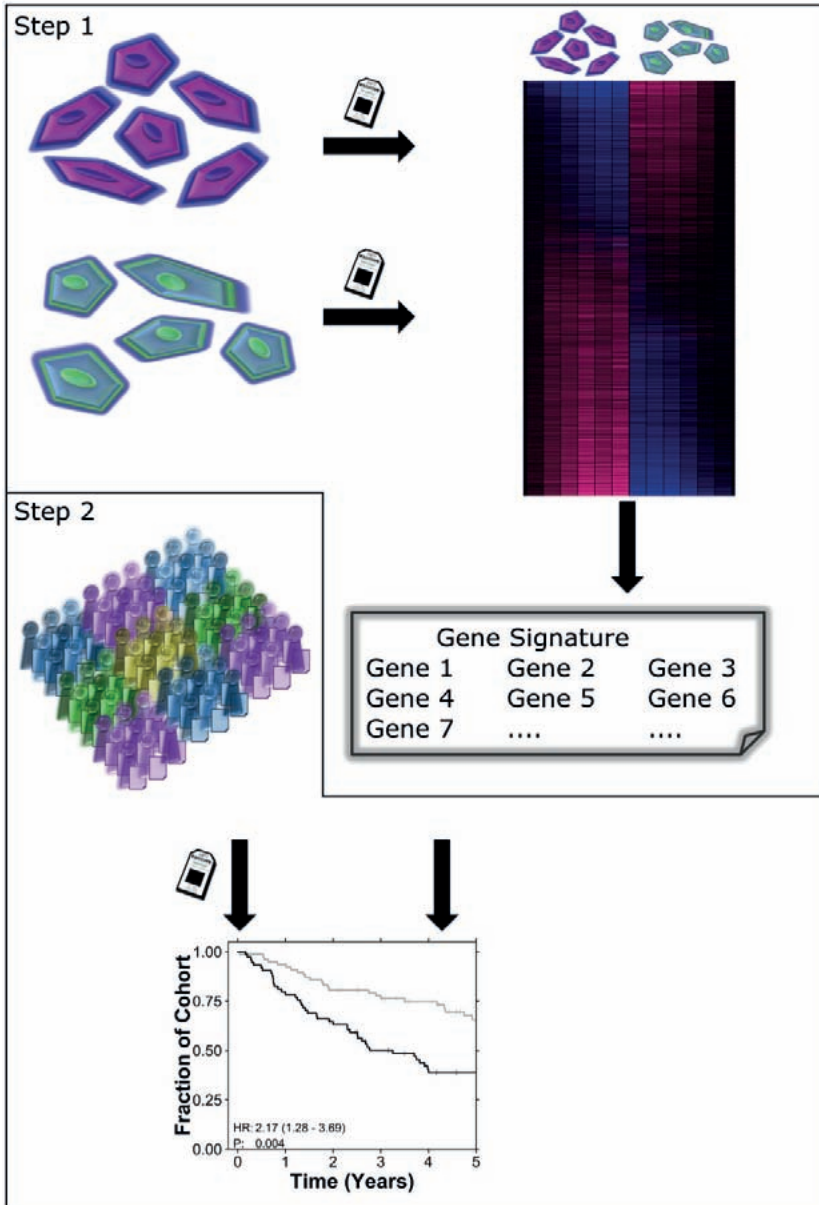


Figure 1.2

Schematic overview of biology-derived gene-expression signature development. In step 1 gene-expression profiling is performed on cells grown under distinct culture conditions *in vitro* to identify differentially expressed genes between the conditions. These genes are then used as signature and tested in patient data to address whether they have prognostic and/or predictive power (step 2).

A microenvironmental feature studied extensively at the gene expression level is hypoxia^{31,50-52}. Series of hypoxia signatures have been created, but only a few were validated in follow-up studies. One of these is the hypoxia metagene³¹, which was derived from a set of well-characterized hypoxia-related genes from which 'seed'-genes were used to identify co-expressed genes in patient data. This larger gene-set, the hypoxia metagene, is subsequently used to subgroup patients with the idea that patients with a high hypoxia metagene score have worse survival. The methodology was refined and validated in follow-up studies^{58,59}.

Challenges in microarray gene expression profiling

Although a large number of patient data-based signatures have been created for specific cancer types and generated from a similar starting point, the overlap between such signatures is extremely small^{27,28,60,61}. The initially reported classifier performance is often very strong, but validation in other datasets often fails^{60,62,63}. This reflects both the significant technical challenges involved in introducing microarrays into a clinical setting, and over-fitting caused by large numbers of genes and relatively small patient cohorts. Taken together, these data also help to explain why only a few markers are currently being tested in large prospective trials.

To make them clinically useful, variability across a large number of parameters has to be taken into account for any gene expression microarray analysis. This variability arises from the biological samples, hybridization protocols, scanning and analysis methods^{60,64}. Differences in the use of these parameters are thus likely to cause discrepancies between signatures. Indeed poor inter-platform correlations have been reported^{65,66} and the use of different statistical procedures in one dataset could identify multiple highly prognostic signatures³³. Since there are multiple microarray platforms, dozens of microarray pre-processing methods and numerous statistical and bioinformatics approaches to create gene expression signatures, standardization is challenging to achieve.

In addition, the number of monitored genes (usually tens of thousands) in gene expression microarrays is large and the number of samples is relatively small. This makes multiple testing a serious issue in this type of research^{60,67}. Without multiple testing correction the danger exists that false prognostics are identified. Corresponding to this, several studies have shown that signature performance can be overestimated^{62,67,68}. Inherent to the data dimensionality issue, modeling studies have demonstrated that larger sample sizes would result in more robust results^{69,70}.

Despite these hurdles, other methods to compare different gene expression markers show consistency, not in gene-wise overlap, but in prognostic performance. Comparison of a range of signatures in both breast and non-small-cell lung cancer revealed that signatures show significant agreement in the identified prognostic groups^{61,71,72}. These data suggest that although signatures disagree in gene content, they track common underlying biology that determine patient outcome. Substantial data indicate that a major contributor to the prognostic power of signatures is proliferation^{71,73-75}. It has been shown that a high number of published signatures contain a proliferation cluster of genes^{72,73}. In a subset of these signatures prognostic performance even improved when only the proliferation genes were taken to classify patients⁷².

Hypothesis and outline of this thesis

As outlined in the previous paragraphs proliferation is one of the major biological processes implicated in the prognostic power of successful signatures. We hypothesized that a pure proliferation signature would have high prognostic power across a range of cancer types and could add value to current clinical practice. **Chapter 2** describes a proliferation signature created from *in vitro* derived gene expression microarray data. This proliferation signature was able to identify groups of patients with poor survival in different datasets spanning several cancer types. We further related expression of the proliferation signature to mitotic index, another proliferation measure, in one of the studied datasets.

Correlating gene expression signatures to biological processes measured with immunohistochemistry (IHC), could serve as a validation step to link biology-driven signatures with other measures of the process they allegedly track and might potentially elucidate the processes driving a patient data-based signature. We propose an *in vivo* based methodology to address these issues in **chapter 3**. In this *in vivo* based model we once more related expression of the proliferation signature to a measure of proliferation.

To investigate the clinical applicability of the proliferation signature, we reduced the number of genes in the signature in order to evaluate it with PCR rather than microarrays. In **chapter 4** this refined proliferation signature was validated, first in two large patient microarray meta-datasets, and consecutively *in vitro*, *in vivo* and in an independent patient dataset with PCR.

The second part of this thesis relates to the high dimensionality of gene expression microarrays and the resulting data analysis issues. We hypothesize that the multiple testing issue is not only crucial in the signature creation phase, but in addition plays an important role in signature testing. Therefore, in **chapter 5** the effect of random generated signatures on prognostic findings is tested in a series of datasets. The percentage of random markers reaching statistical significance reached dramatic levels, but varied widely dependent on the dataset. We propose a simple method to test the likelihood that a signature performs better than chance.

A recent review on gene expression-based classifiers in non-small-cell lung cancer⁶³ judged that published markers so-far had serious problems in their design and analyses, rendering them immature for clinical applicability. Several microarray studies were scored on the basis of a predefined set of criteria for study design, reporting and analyses. Each study either passed or failed a criterion, which is a rather black and white evaluation. For example a study providing three out of four specified clinical variables failed that criterion similar to a study that provided no covariates. In addition a couple of gene expression-based markers were tested in a new independent patient dataset. Data pre-processing as described in the original studies was not followed and the classifiers did not validate. In **chapter 6** we argue that results in this review are not totally presented in a complete and accurate way leading to unnecessary harsh and biased conclusions.

In **chapter 7** we sought to determine the influence of microarray pre-processing methodology on signature performance. We hypothesized that differences in pre-processing procedure could lead to erroneous conclusions regarding signature validation. A previously published gene expression classifier for non-small-cell lung cancer³³ was tested in a large patient dataset³² pre-processed with 24 different schemes. Prognosticity was highly dependent on the preprocessing schedule used, emphasizing the importance of consistency in data analyses. Furthermore we validated two previously published non-small-cell lung cancer markers^{27,33} this new dataset.

In conclusion, we created a highly prognostic gene expression-based marker for proliferation. This marker was successfully translated to a PCR-based classifier, which will increase clinical applicability. We further developed and validated two novel methods for assessing the robustness of gene expression-based biomarkers. Taking these into account will improve reported classifier validity and lead to a faster translation to a clinical setting.

References

1. World Health Organization; Cancer - Key Facts 2007.
2. Garcia, M., *et al.* Global Cancer Facts & Figures 2007. *Atlanta, GA: American Cancer Society* (2007).
3. http://www.ikcnet.nl/uploaded/docs/Landelijk/cijfers/incidentie%202008/A01_NL.xls.
4. http://www.ikcnet.nl/uploaded/docs/Landelijk/cijfers/incidentie%202008/B1_NL.xls.
5. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
6. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
7. Hahn, W.C., *et al.* Creation of human tumour cells with defined genetic elements. *Nature* **400**, 464-468 (1999).
8. Rand, V., *et al.* Sequence survey of receptor tyrosine kinases reveals mutations in glioblastomas. *Proc Natl Acad Sci U S A* **102**, 14344-14349 (2005).
9. Greenman, C., *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-158 (2007).
10. Dalgliesh, G.L., *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**, 360-363 (2010).
11. Jain, R.K. Barriers to drug delivery in solid tumors. *Sci Am* **271**, 58-65 (1994).
12. Padera, T.P., *et al.* Pathology: cancer cells compress intratumour vessels. *Nature* **427**, 695 (2004).
13. Wouters, B.G. & Koritzinsky, M. Hypoxia signalling through mTOR and the unfolded protein response in cancer. *Nat Rev Cancer* **8**, 851-864 (2008).
14. Dewhirst, M.W., Cao, Y. & Moeller, B. Cycling hypoxia and free radicals regulate angiogenesis and radiotherapy response. *Nat Rev Cancer* **8**, 425-437 (2008).
15. Milosevic, M., Fyles, A., Hedley, D. & Hill, R. The human tumor microenvironment: invasive (needle) measurement of oxygen and interstitial fluid pressure. *Semin Radiat Oncol* **14**, 249-258 (2004).
16. Nordmark, M., *et al.* Differential risk assessments from five hypoxia specific assays: The basis for biologically adapted individualized radiotherapy in advanced head and neck cancer patients. *Radiother Oncol* **83**, 389-397 (2007).
17. Leo, C., Giaccia, A.J. & Denko, N.C. The hypoxic tumor microenvironment and gene expression. *Semin Radiat Oncol* **14**, 207-214 (2004).
18. Perou, C.M., *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747-752 (2000).
19. Ramaswamy, S., *et al.* Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci U S A* **98**, 15149-15154 (2001).
20. Gordon, G.J., *et al.* Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. *Cancer Res* **62**, 4963-4967 (2002).
21. van 't Veer, L.J., *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530-536 (2002).
22. Chang, H.Y., *et al.* Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* **2**, E7 (2004).
23. Paik, S., *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* **351**, 2817-2826 (2004).
24. Carter, S.L., Eklund, A.C., Kohane, I.S., Harris, L.N. & Szallasi, Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* **38**, 1043-1048 (2006).

25. Chung, C.H., *et al.* Increased epidermal growth factor receptor gene copy number is associated with poor prognosis in head and neck squamous cell carcinomas. *J Clin Oncol* **24**, 4170-4176 (2006).
26. Bild, A.H., *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353-357 (2006).
27. Lau, S.K., *et al.* Three-gene prognostic classifier for early-stage non small-cell lung cancer. *J Clin Oncol* **25**, 5562-5569 (2007).
28. Chen, H.Y., *et al.* A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* **356**, 11-20 (2007).
29. Liu, R., *et al.* The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* **356**, 217-226 (2007).
30. Jones, J. & Libermann, T.A. Genomics of renal cell cancer: the biology behind and the therapy ahead. *Clin Cancer Res* **13**, 685s-692s (2007).
31. Winter, S.C., *et al.* Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. *Cancer Res* **67**, 3441-3449 (2007).
32. Shedden, K., *et al.* Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* **14**, 822-827 (2008).
33. Boutros, P.C., *et al.* Prognostic gene signatures for non-small-cell lung cancer. *Proc Natl Acad Sci U S A* **106**, 2824-2828 (2009).
34. van de Vijver, M.J., *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347**, 1999-2009 (2002).
35. Buyse, M., *et al.* Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* **98**, 1183-1192 (2006).
36. Mook, S., *et al.* The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1-3 positive lymph nodes in an independent validation study. *Breast Cancer Res Treat* **116**, 295-302 (2009).
37. Bueno-de-Mesquita, J.M., *et al.* Validation of 70-gene prognosis signature in node-negative breast cancer. *Breast Cancer Res Treat* **117**, 483-495 (2009).
38. Bogaerts, J., *et al.* Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. *Nat Clin Pract Oncol* **3**, 540-551 (2006).
39. Cardoso, F., *et al.* Clinical application of the 70-gene profile: the MINDACT trial. *J Clin Oncol* **26**, 729-735 (2008).
40. Sparano, J.A. TAILORx: trial assigning individualized options for treatment (Rx). *Clin Breast Cancer* **7**, 347-350 (2006).
41. Paik, S., *et al.* Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* **24**, 3726-3734 (2006).
42. Toi, M., *et al.* Clinical significance of the 21-gene signature (Oncotype DX) in hormone receptor-positive early stage primary breast cancer in the Japanese population. *Cancer* **116**, 3112-3118 (2010).
43. Habel, L.A., *et al.* A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res* **8**, R25 (2006).
44. Sorlie, T., *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**, 10869-10874 (2001).
45. Sorlie, T., *et al.* Repeated observation of breast tumor subtypes in independent gene expression data

- sets. *Proc Natl Acad Sci U S A* **100**, 8418-8423 (2003).
46. Sotiriou, C., *et al.* Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* **100**, 10393-10398 (2003).
 47. Hu, Z., *et al.* The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* **7**, 96 (2006).
 48. Torres-Roca, J.F., *et al.* Prediction of radiation sensitivity using a gene expression classifier. *Cancer Res* **65**, 7169-7176 (2005).
 49. Chen, J.L., *et al.* The genomic analysis of lactic acidosis and acidosis response in human cancers. *PLoS Genet* **4**, e1000293 (2008).
 50. Chi, J.T., *et al.* Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. *PLoS Med* **3**, e47 (2006).
 51. Sung, F.L., *et al.* Genome-wide expression analysis using microarray identified complex signaling pathways modulated by hypoxia in nasopharyngeal carcinoma. *Cancer Lett* **253**, 74-88 (2007).
 52. Seigneuric, R., *et al.* Impact of supervised gene signatures of early hypoxia on patient survival. *Radiother Oncol* **83**, 374-382 (2007).
 53. Bissell, M.J. & Radisky, D. Putting tumours in context. *Nat Rev Cancer* **1**, 46-54 (2001).
 54. Pedersen, T.X., *et al.* Laser capture microdissection-based in vivo genomic profiling of wound keratinocytes identifies similarities and differences to squamous cell carcinoma. *Oncogene* **22**, 3964-3976 (2003).
 55. Iyer, V.R., *et al.* The transcriptional program in the response of human fibroblasts to serum. *Science* **283**, 83-87 (1999).
 56. Chang, H.Y., *et al.* Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* **102**, 3738-3743 (2005).
 57. Zhao, H., *et al.* Alteration of gene expression signatures of cortical differentiation and wound response in lethal clear cell renal cell carcinomas. *PLoS One* **4**, e6039 (2009).
 58. Buffa, F.M., Harris, A.L., West, C.M. & Miller, C.J. Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *Br J Cancer* **102**, 428-435 (2010).
 59. Ghazoui, Z., *et al.* Close and stable relationship between proliferation and a hypoxia metagene in aromatase inhibitor treated ER-positive breast cancer. *Clin Cancer Res* (2011).
 60. Dupuy, A. & Simon, R.M. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* **99**, 147-157 (2007).
 61. Fan, C., *et al.* Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* **355**, 560-569 (2006).
 62. Michiels, S., Koscielny, S. & Hill, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* **365**, 488-492 (2005).
 63. Subramanian, J. & Simon, R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Natl Cancer Inst* **102**, 464-474 (2010).
 64. Miller, L.D., *et al.* Optimal gene expression analysis by microarrays. *Cancer Cell* **2**, 353-361 (2002).
 65. Kuo, W.P., Jenssen, T.K., Butte, A.J., Ohno-Machado, L. & Kohane, I.S. Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* **18**, 405-412 (2002).
 66. Tan, P.K., *et al.* Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* **31**, 5676-5684 (2003).
 67. Ntzani, E.E. & Ioannidis, J.P. Predictive ability of DNA microarrays for cancer outcomes and

- correlates: an empirical assessment. *Lancet* **362**, 1439-1444 (2003).
68. Ein-Dor, L., Kela, I., Getz, G., Givol, D. & Domany, E. Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics* **21**, 171-178 (2005).
 69. Zien, A., Fluck, J., Zimmer, R. & Lengauer, T. Microarrays: how many do you need? *J Comput Biol* **10**, 653-667 (2003).
 70. Ein-Dor, L., Zuk, O. & Domany, E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci U S A* **103**, 5923-5928 (2006).
 71. Haibe-Kains, B., Desmedt, C., Sotiriou, C. & Bontempi, G. A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? *Bioinformatics* **24**, 2200-2208 (2008).
 72. Wirapati, P., *et al.* Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* **10**, R65 (2008).
 73. Whitfield, M.L., George, L.K., Grant, G.D. & Perou, C.M. Common markers of proliferation. *Nat Rev Cancer* **6**, 99-106 (2006).
 74. Desmedt, C., *et al.* Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res* **14**, 5158-5165 (2008).
 75. Weigelt, B., Baehner, F.L. & Reis-Filho, J.S. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. *J Pathol* **220**, 263-280 (2010).

Chapter 2

Robust prognostic value of a knowledge-based proliferation signature across large patient microarray studies spanning different cancer types

Published in: British Journal of Cancer 2008; 99(11): 1884–1890

Maud H.W. Starmans, Balaji Krishnapuram, Harald Steck, Hugo Horlings, Dimitry S.A. Nuyten, Marc J. van de Vijver, Renaud Seigneuric, Francesca M. Buffa, Adrian L. Harris, Bradly G. Wouters, Philippe Lambin

Abstract

Tumour proliferation is one of the main biological phenotypes limiting cure in oncology. Extensive research is being performed to unravel the key-players in this process. To exploit the potential of published gene expression data, creation of a signature for proliferation can provide valuable information on tumour status, prognosis and prediction. This will help individualizing treatment and should result in better tumour control, and more rapid and cost-effective research and development.

From *in vitro* published microarray studies, two proliferation signatures were compiled. The prognostic value of these signatures was tested in five large clinical microarray datasets. More than 1,000 patients with breast, renal or lung cancer were included. One of the signatures (110 genes) had significant prognostic value in all datasets. Stratifying patients in groups resulted in a clear difference in survival (p-values < 0.05). Multivariate Cox-regression analyses showed that this signature added substantial value to clinical factors used for prognosis. Further patient stratification was compared to patient stratification with several well-known published signatures. Contingency tables and Cramer's V statistics indicated that these primarily identify the same patients as the proliferation signature does.

The proliferation signature is a strong prognostic factor, with the potential to be converted into a predictive test. Further, evidence is provided that supports the idea that many published signatures track the same biological processes and that proliferation is one of them.

Introduction

The ability to predict outcome and to identify key-players in biological mechanisms that lead to poor outcome, are two important objectives in cancer research. Recently, efforts to exploit gene expression profiling have been made to identify gene sets, or so-called gene signatures, that can improve diagnosis and risk stratification¹. A drawback of most of the studies performed is that supervised analysis methods are utilized to acquire such signatures. In this approach, patient microarray and clinical data are used to find gene sets that correlate with tumour type or survival. This often results in gene sets with a very high prognostic value in the studied dataset. However, comparative testing of these signatures in other patient datasets has been limited, and the overlap in selected genes of different comparable studies is small². If such a signature can be applied to other datasets it may well be restricted to a certain patient population and cancer type. In addition, the gene sets obtained with this method are often difficult to interpret with respect to the underlying biological mechanism^{3,4}. Further Dupuy and Simon⁵, showed in a recent review that many of these studies show flaws in methodology.

An alternative approach to identify prognostic signatures is based on defining gene sets involved in a biological process or specific environmental condition that is suspected of influencing treatment response or patient outcome. In this approach, *in vitro* gene expression profiling is used to identify gene sets that play an important role in a specific biological process. The identified gene set is then applied to gene expression data from patients to evaluate its prognostic value. This approach has a more broad application because the gene sets can be applied in almost every patient group. Furthermore, it can be used not only to investigate whether a certain process is important in a distinct cancer type or patient group but also potentially to select patients in those groups that would be expected to benefit from therapies directed to the biological process of interest¹. Examples of gene sets attained with this approach are the wound⁶, hypoxia⁷⁻⁹ and “invasiveness” (IGS)¹⁰ signatures. These studies show that the deduced signatures can be used for risk stratification in very different types of cancers^{6,7,9,10}, presumably because of common core pathways that are influencing outcome in these diverse clinical groups. Another potential benefit of this unsupervised approach is that it can potentially identify the functional regulators within a signature that drive the studied process¹¹ and thus reveal new targeting candidates. Recently Fan *et al*¹², compared the performance of

several supervised and unsupervised derived gene sets and found that both types of signatures showed high concordance in prognostic power¹³.

One of the biological processes often implicated in gene expression signatures is cell proliferation. The rate of tumour cell proliferation is a major contributor to treatment response with both chemotherapy and radiotherapy¹⁴ and is reflected in the fact that overall treatment time (e.g. duration of radiotherapy) is an important contributor to outcome¹⁵. In a recent review Whitfield *et al*¹⁶, showed that proliferation may underlie the prognostic power of many previously identified signatures. He showed that in almost every supervised derived signature a large subset of genes involved in proliferation is included^{3,17,18}. In some cases, these classifiers have even been designated as ‘proliferation’ signatures although their derivation was not based on this phenotype. Two of these signatures have recently made it to the clinical setting as a diagnostic tool for patients with breast cancer^{19,20}. Based on these results, we hypothesized that derivation of a specific *in vitro* derived signature based solely on proliferation may provide valuable information on tumour status, prognosis and outcome prediction across diverse tumour types.

Materials and methods

Datasets

Patient microarray and clinical follow-up data were collated to test the clinical value of the signatures. Datasets are publicly available in the microarray databases Gene Expression Omnibus (GEO) and Stanford Microarray Database (SMD) or elsewhere. Accessory clinical and follow-up data were also given or provided by the authors on request. In Table 2.1 an overview of the datasets and where they are accessible is provided. Data filtering and pre-processing is explained in the supplementary data (Supplementary materials and methods). Datasets were imported in Matlab (Matlab 7.1, The Mathworks, Massachusetts, USA). Unless indicated otherwise, analyses were performed in this program.

Signature score calculation

Expression data of the genes in the signature was extracted from the dataset. The following step was used to calculate a signature score for each patient in the dataset. This score was defined as the weighted average expression value of the genes in the signature (Equation 1). A weight of -1 or 1 was assigned to each gene,

dependent on the phenotype the gene represented. Weight assignment is described in the results and supplementary data (Supplementary materials and methods).

The signature score then reflects the status of the studied process in a tumour. When a gene was represented by more than one probe on an array, the expression of the probes was averaged before signature calculation.

$$\text{Score} = \frac{\sum_{i=1}^N w_i \cdot \text{exp}_i}{N} \quad \text{Equation (1)}$$

With: Score = signature score
N = number of genes in the signature
i = gene
 w_i = weight of gene i
 exp_i = gene expression of gene i

Statistical analysis

A loop of 1,000 clustering repeats with the K-means clustering function in Matlab was applied to split the patients in two groups according to their signature score. Outcome in the two groups was analyzed and compared by the Kaplan-Meier method. Differences in outcome were tested for statistical significance by the log-rank test for different common end-points. For breast and renal cancer the common end-points are 5-years and 10-years survival, for lung cancer these are 2-years and 5-years survival, all end-points were analyzed when follow-up was long enough. Results for the log-rank tests are given as the average, standard deviation and the range of the p-values, also the percentage of p-values from the 1,000 clustering runs that were significant was calculated to evaluate the prognostic power of the signature and stability of the clustering.

Multivariate Cox regression analysis with stepwise backward selection procedure was performed in SPSS (SPSS 12.0.1, SPSS Inc, Illinois, USA) to show the clinical relevance of the proliferation signature.

AUC model calculation

Matlab was used to integrate all parameters in a model with and without addition of the signature to the clinical parameters. Differences between the models were assessed using receiver operator curve (ROC) analysis by calculating the AUC.

Further details are provided in the supplementary data (Supplementary materials and methods).

Random signature testing

A method to test a predefined number of random signatures of a predefined size on all the datasets was developed. To show the strength of a signature 10,000 random generated gene sets, with sizes equal to the size of the signature of interest, were tested on the datasets. These random gene sets were tested in a similar manner as the other signatures.

Mitotic index scoring

MI was assessed (as part of histological grading) in the 295 breast tumours of the van de Vijver dataset¹³ using a microscope with a field diameter of 0.44 mm using a x40 objective. The area with the highest mitotic activity was selected and mitotic figures were counted in 10 consecutive fields. Tumours were assigned to the following groups based on the mitotic counts:

- Group 1: 0-5 mitoses in 10 high power fields
- Group 2: 6 -10 mitoses in 10 high power fields
- Group 3: ≥ 11 mitoses in 10 high power fields

Table 2.1 A Overview of the analyzed patient microarray datasets, B: Number of signature genes represented in the microarray dataset (number of gene identifiers on the arrays are given between brackets).

| A | | | | |
|----------------------|--------------------|--------------------|---|----------------------|
| Dataset | Cancer site | # patients | Source | |
| Miller | Breast | 251 | GEO accession GSE3494: http://www.ncbi.nlm.nih.gov/projects/geo/ | |
| Wang | Breast | 286 | GEO accession GSE2034: http://www.ncbi.nlm.nih.gov/projects/geo/ | |
| van de Vijver | Breast | 295 | http://microarray-pubs.stanford.edu/wound_NKI/ | |
| Zhao | Renal | 177 | SMD: http://smd.stanford.edu/ | |
| Beer | Lung | 86 | http://dot.ped.med.umich.edu:2000/ourimage/pub/Lung/index.html | |
| B | | | | |
| Dataset | Signature 1 | Signature 2 | Wound signature | IGS signature |
| Miller | 455 (1120) | 104 (228) | 415 (1030) | 176 (516) |
| Wang | 350 (667) | 87 (158) | 346 (614) | 131 (270) |
| van de Vijver | 192 (242) | 51 (59) | 171 (195) | 67 (87) |
| Zhao | 257 (415) | 47 (82) | 280 (446) | 83 (132) |
| Beer | 192 (224) | 45 (51) | 171 (195) | 63 (76) |

Results

Signature derivation

From published microarray studies two different proliferation signatures were compiled. Whitfield *et al*²¹, studied the cell cycle in HeLa cells (cervix cancer cell line). Microarrays were performed on synchronized cell cultures at different time-points and genes that showed a periodic variation were selected. These genes were grouped according to the cell cycle phase in which their expression peaked. We propose that this gene set could be employed as a specific proliferation signature.

Another method to derive a proliferation signature with microarrays was employed by Chang *et al*⁶. Human fibroblasts were serum starved for 48 hours and then stimulated with serum to simulate a wound response. One of the most consistent and important effects in the serum response program is stimulation of proliferation. Abnormal proliferation is also a consistent characteristic of cancer cells, irrespective of a wound response⁶. Chang *et al*⁶, therefore discarded the genes with a periodic behavior to specifically study the wound response. Here we propose that the set of genes discarded from the wound signature is a good representation of a proliferation signature. This signature is a subset of the signature derived from Whitfield *et al*²¹, however we postulate that it is a better representative of proliferation and will be a better prognostic factor, since only this gene set shows a change in expression upon serum stimulation.

The wound⁶ and IGS¹⁰ signature are two promising published unsupervised derived signatures. Further the second proliferation signature is derived from the same *in vitro* data as the wound signature. Therefore these signatures were also analyzed.

Comparison of two proliferation signatures

Signature 1²¹ and signature 2⁶ consist of respectively 1,134 and 199 cloneIDs that map to 815 and 154 unique UnigenesIDs, respectively. The distribution of genes in the different cell cycle phases for the two signatures is distinct (Supplementary data Table S2.1), indicating that the signatures are different. Signature 1 shows equal proportions of genes in the defined cell cycle phases. However in signature 2 more genes are involved in G₂ and clearly less genes are involved in M/G₁.

Outcome prediction with proliferation signatures

The signatures were tested for their clinical relevance on several publicly available microarray datasets (Table 2.1). Signatures were evaluated using a signature score (Equation 1), which is defined as a weighted average of the expression of the genes in the signature. To calculate the signature score, weights were defined for each gene. After translating the signatures into UnigeneIDs (build199) and weight assignment several genes were discarded from analyses, for these genes weight assignment was ambiguous, details are provided in the supplementary materials and methods. The final signatures consist of respectively 508 and 110 UnigeneIDs for signature 1 and 2.

In every dataset a signature score (Equation 1) was calculated for each patient. The patients were separated in two groups by clustering these signature scores, to obtain a natural separation rather than using an arbitrary value like the median to split the patients. This clustering was repeated 1,000 times to assess the stability of the group assignment. Results of the log-rank tests are given in supplementary data Table S2.2 and in Figure 2.1 the Kaplan-Meier curves for signature 2 are shown. Signature 2 gives clear risk stratification in all datasets, all p-values of the 1,000 clustering runs < 0.05 . Results of the log-rank test show not only that signature 2 gives a better risk stratification than signature 1, also the overall robustness of the separation is stronger, indicated by the small standard deviations. Nevertheless both signatures show very good prognostic value on the three breast cancer datasets. The range and standard deviations of the 1,000 clustering runs also show that the results are robust for these datasets and that splitting patients based on clustering of signature scores is stable.

Statistical analysis of signature scores

Multivariate Cox-regression analyses were performed to investigate whether the association between the best proliferation signature and outcome was independent of clinical prognostic factors. The variables analyzed differed per dataset, since different clinical factors are provided (Supplementary data Table S2.3). A stepwise backward selection procedure was performed to select the variables that are prognostic factors; the end-point is 10-years for breast and renal cancer and 5-years for lung cancer. Follow-up time in the Wang *et al*²² dataset is not long enough, in that dataset 5-years was used. In Table 2.2 the factors selected with this procedure are given for all the datasets, choosing another end-point did not influence the

results dramatically (Supplementary data Table S2.4). In 4 out of 5 datasets the proliferation signature is included in the model as a prognostic factor of outcome. In 3 datasets this was highly significant and in the fourth it reached border significance.

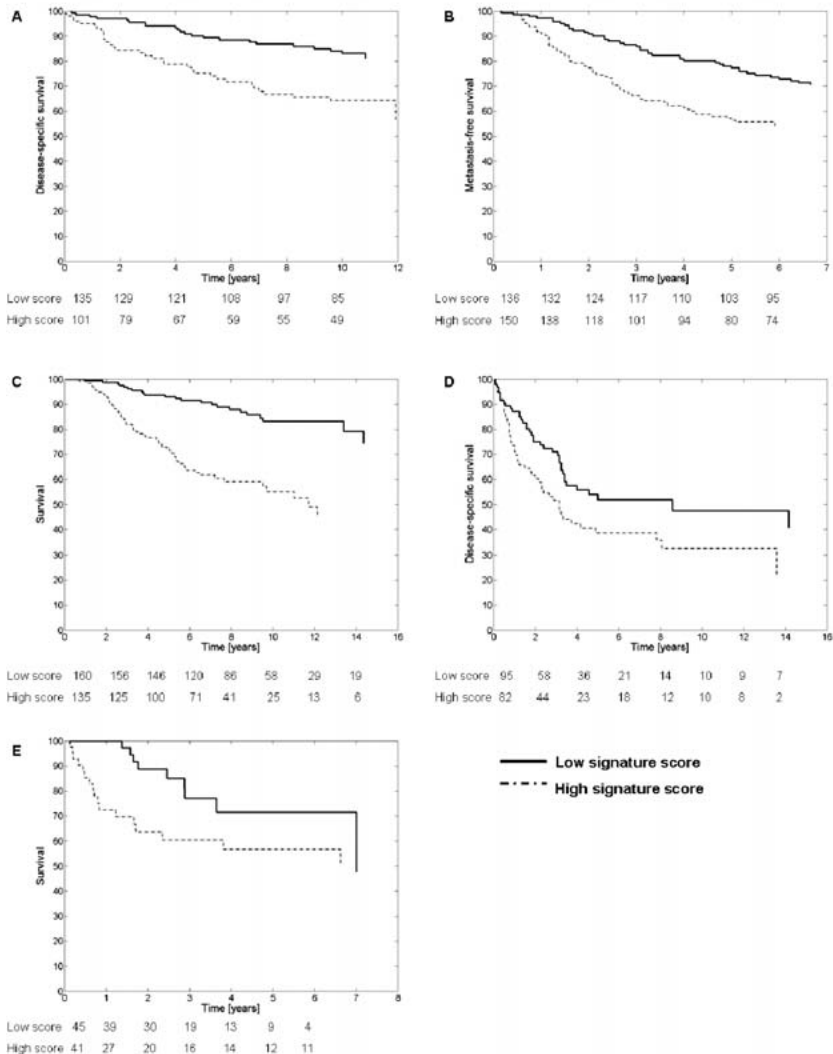


Figure 2.1

A signature score was calculated for each patient in the different datasets. These scores were used to cluster the patients in two groups, one with low expression and one with high expression of the signature. Kaplan-Meier survival curves for the two groups were compared (A: Miller dataset, B: Wang dataset, C: van de Vijver dataset, D: Zhao dataset, E: Beer dataset).

Table 2.2 Clinical parameters selected with stepwise backward selection in multivariate Cox-regression analyses including signature 2.

| Miller | Hazard ratio (95% CI) | p-value |
|----------------------|------------------------------|----------------|
| Tumor size | 3.3 (1.7 - 6.6) | 0.0006 |
| LNS† | 2.8 (1.6 - 5.0) | 0.0003 |
| Proliferation** | 3.4 (1.4 - 8.2) | 0.0052 |
| Wang | Hazard ratio (95% CI) | p-value |
| Proliferation** | 2.6 (1.5 - 4.4) | 0.0004 |
| van de Vijver | Hazard ratio (95% CI) | p-value |
| Age | 0.95 (0.91 - 0.99) | 0.0096 |
| Tumor size* | 1.5 (0.93 - 2.5) | 0.0962 |
| Elston grade | 2.2 (1.4 - 3.4) | 0.0003 |
| Proliferation** | 21 (1.8 - 234) | 0.0148 |
| Zhao | Hazard ratio (95% CI) | p-value |
| Performance status | 1.3 (1.1 - 1.6) | 0.0069 |
| Grade | 1.5 (1.0 - 2.1) | 0.0260 |
| Stage | 3.3 (2.5 - 4.4) | < 0.0001 |
| Beer | Hazard ratio (95% CI) | p-value |
| Age | 1.0 (1.0 - 1.1) | 0.0338 |
| Stage | 2.3 (1.5 - 3.5) | 0.0002 |
| Proliferation** | 1.8 (0.9 - 3.5) | 0.0884 |

* Categories: ≤ 2 cm or > 2 cm

† LNS: lymph-node status

** Proliferation: proliferation signature 2

AUC calculations

The area under the receiver-operator curve (AUC) was calculated for each clinical parameter and the best proliferation signature. Results of this analyses show that the proliferation signature has a high AUC compared to the clinical parameters in all datasets (Supplementary data Table S2.3).

To quantify the gain in prognostic power obtained with this signature a model of the clinical factors with and without the signature was generated and evaluated with the AUC. Part of the dataset was used as training set, to generate the model, and the other part as a test set. Only the datasets with more than 1 clinical parameter and more than 150 patients are included. Different sizes of training and test sets were evaluated; the overall performance did not change significantly (data not shown). The results shown in Figure 2.2 were produced with 150 and 100 samples as training set for the breast cancer and the renal cancer datasets respectively. In 2 out of 3 datasets the AUC increased significantly when the

proliferation signature was added to the model (Figure 2.2, p-values paired t-test $\ll 0.0001$).

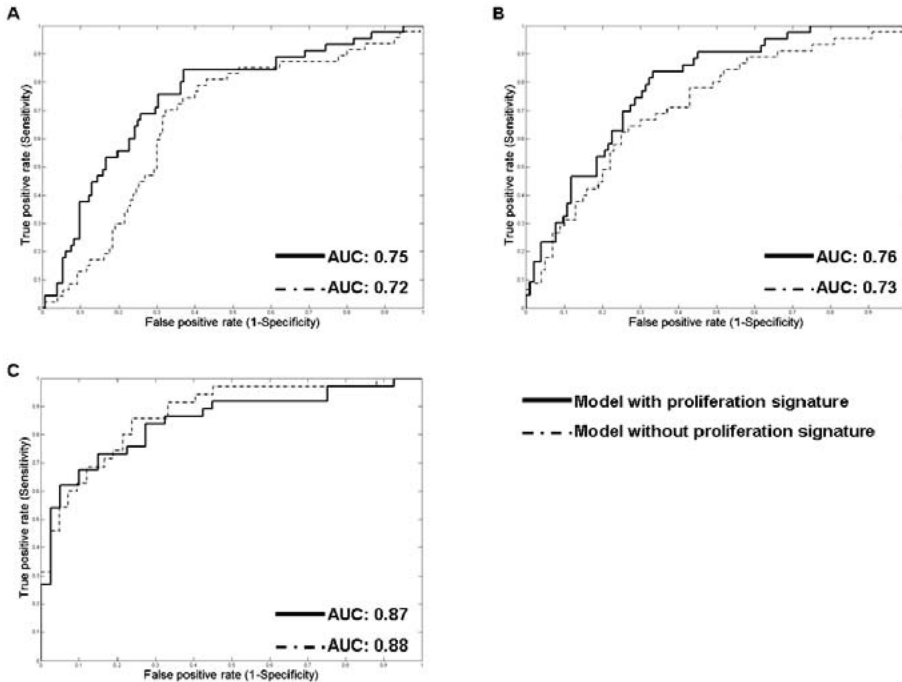


Figure 2.2

A model of the clinical factors with and without the signature was generated. Receiver operator curves (ROC) were used to compare the two models in three datasets. (A: Miller dataset, B: van de Vijver dataset, C: Zhao dataset).

Random signature testing

In order to show the strength of the proliferation signature, 10,000 random generated signatures were tested on all datasets. Of these 10,000 no signature gave a significant result on all datasets.

Comparison to other signatures

Log-rank tests and Kaplan-Meier survival curves show that the wound and IGS signature give clear risk stratification in 4 and 5 datasets respectively (Supplementary data Table S2.2). Further inclusion of these signatures in multivariate Cox-regression analyses (Supplementary data Table S2.4) gives the indication that combining signatures, like the proliferation and IGS signature in one

of the datasets can strengthen the prognostic power of microarray profiling in a clinical setting. Combining the proliferation and wound signature did not add value in any of the datasets.

To investigate whether different signatures identify the same patients, two-way contingency table analyses (Supplementary materials and methods) were performed to compare the patient classification of the proliferation signature to the classification of other signatures. For the gene sets identified in three of the five datasets^{13,23,24}, group classification was available (details are provided in the Supplementary materials and methods), these and the wound and IGS signature were evaluated. Contingency table analyses and Cramer's V statistics (Supplementary data Tables S2.5-S2.9) show that the proliferation signature has a strong association with all other signatures evaluated, indicating that these signatures identify the same patients.

Proliferation signature validation

The proliferation signature is extracted from *in vitro* data, however this does not necessarily mean that the signature truly tracks proliferation *in vivo*. To investigate this mitotic index (MI) was assessed for the van de Vijver¹³ dataset, which was scored in three classes. In Figure 2.3 a boxplot is shown of the proliferation score versus the MI. There is a clear correlation between the average proliferation signature score and the three classes of MI (correlation coefficient: 0.968).

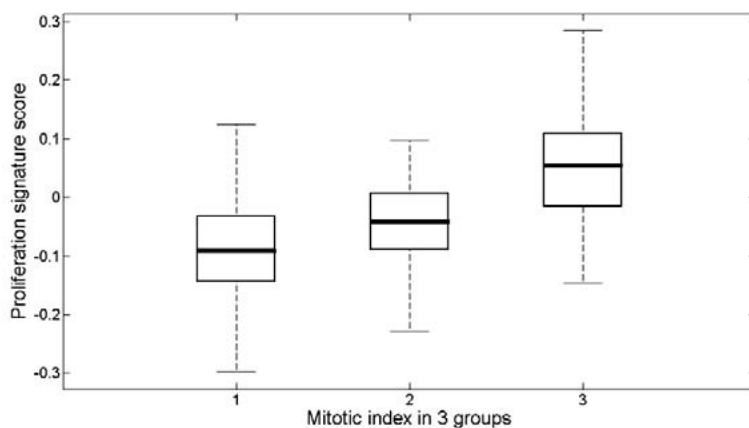


Figure 2.3

Correlation between the proliferation signature score and the mitotic index in the van de Vijver dataset.

Discussion

We derived a proliferation signature from *in vitro* microarray studies based only on genes that differ in expression in different parts of the cell cycle^{6,21}. Results show that the proliferation signature has a high value in patient risk stratification in five large clinical studies involving more than 1,000 patients and three different cancer sites. This contrasts with previous studies that rarely validate signatures in more than one large independent dataset.

Our data indicate that the proliferation signature can be combined with other phenotype based signatures, to further improve patient stratification. The fact that large clusters of proliferation genes are identified in many gene signatures^{3,10,17,18,25-27} raises the possibility that many of previously reported gene signatures, including the wound signature, may be highly influenced by proliferation. Fan *et al*¹², has previously suggested that many signatures track a common set of biologic phenotypes and have therefore a similar prognostic strength. Whitfield *et al*¹⁶, has further suggested that one of these processes is proliferation. The performance of the proliferation signature in our study supports this idea. Comparisons of the proliferation signature to five other gene sets showed that these signatures primarily identify the same patients as the proliferation signature.

Some reports refer to their signature as a proliferation signature^{3,27}. However in these supervised studies not all genes in the signature are related to proliferation and therefore cannot be referred to strictly as general proliferation signatures. For example, Dai *et al*³, used a supervised approach to determine a signature associated with metastasis. Many of the identified genes were known to be involved in cell cycle regulation and these authors thus referred to their classifier as a proliferation signature. However only 17 out of 50 genes in this signature are cell cycle related when assessed in the initial gene list of Whitfield *et al*²¹. The same applies to the study of Rosenwald *et al*²⁷, only 28 of the 48 genes that were associated with length of survival are related to proliferation.

The proliferation signature has a high prognostic power, like many signatures, however it is one of the few signatures that may also have a predictive value. It can possibly be used to prescribe a treatment targeting tumour proliferation. Studies indicate that fast proliferating tumours can benefit from accelerated radiotherapy or chemo-radiotherapy^{28,29}. The proliferation signature could possibly be used as basis for a predictive test for patient selection for these treatments.

Previous studies have tried to assess the predictive value of proliferation by means of MI, Ki67 staining and potential doubling time (Tpot) calculation. Overall results of these single-parameter indicators are disappointing³⁰. MI and Ki67 staining are the most promising parameters; however results for these markers are controversial³⁰⁻³³. This can be due to the large chance of misclassification with these single-parameter indicators^{16,33}. Application of multi-parameter indicators, like the proliferation signature, is therefore a more attractive method¹⁶. The proliferation signature shows a clear correlation with MI in one of the tested datasets.

In conclusion we have shown that the application of phenotype based signatures like the proliferation signature can be used in patient risk stratification, in addition to clinical parameters. It has a high prognostic value and unlike other signatures it has the potential to be converted into a predictive test. Furthermore, we provide evidence that supports the idea that many published signatures track the same biological processes and that proliferation is one of them. Whether the proliferation signature can be converted into a predictive test should be evaluated in a large prospective trial in which other measures for proliferation should also be evaluated.

Acknowledgements

Research was supported by Keyzer fonds and Siemens Oncology Care Systems (OCS).

Supplementary material and methods

Signature processing

For all signatures the gene identifiers were translated into UnigeneIDs (Build199) with Source (<http://smd.stanford.edu/>) or Affymetrix data files (www.affymetrix.com: Affx annotation files). After this translation several genes in the proliferation signatures were represented by more than one cloneID. In case these cloneIDs represented the same proliferation status they were included in the signature. However when multiple cloneIDs representing one gene corresponded to different proliferation conditions these genes were discarded. This was approximately 3% of the genes in each signature.

Data filtering and pre-processing

Datasets downloaded from the SMD³⁴ were filtered according to the parameters in the paper. ClonelDs were chosen as gene annotation and the data obtained was log-transformed. For the normalized affymetrix arrays^{22,24} the genes were log-transformed. The Beer *et al*²³ dataset was already preprocessed therefore to perform log-transformation all expression values below 1.1 were set to 1.1, this was similar to the processing performed by Chen *et al*². In all other cases the data was kept in the downloaded format¹³, which was already log-transformed. ClonelDs and affymetrix probeIDs were translated into UnigeneIDs (Build199).

Weight assignment

Genes with a peak expression in S, G₂ and M phase clearly are proliferation genes. Higher expression of these genes will correlate with a larger proportion of cells in S, G₂ and M phase, which indicates more proliferation. That does not apply to genes with a peak expression in G₁, since a G₁ cell may or may not commit to cell cycle. Therefore genes with a peak expression in G₁, 34% and 25% of the genes in signature 1 and 2 respectively, were omitted from further analyses. Equal weights were assigned to the remaining genes in the two proliferation signatures.

For the wound signature weights of -1 and 1 were assigned to genes representing a quiescent and activated wound response respectively. For the IGS signature log ratios were provided in the supplementary data of the paper, a weight of 1 is assigned to genes with a positive log ratio and a weight of -1 to genes with a negative log ratio.

AUC model calculation

Matlab (Matlab 7.1, The Mathworks, Massachusetts, USA) was used to integrate all parameters in a model and evaluate the area under the curve (AUC) of the model with and without addition of the signature to the clinical parameters. All clinical parameters were transformed to numbers, to be able to incorporate them in Matlab, e.g. negative and positive ER-status were set to 0 and 1 respectively. These parameters were incorporated in a model with the classify function of Matlab, which used the diaglinear method. Part of the dataset was used as training set and the other part as a test set. Assignment of samples to test and training set was done at random and repeated 1,000 times.

Contingency table analyses

Contingency tables were used to compare patient classification of the proliferation signature to the patient classification of other gene signatures. For three datasets^{13,23,24} we were able to find the group classification of the gene signatures identified in these studies: the 32-gene p53 signature²⁴, the 70-gene signature¹³ and the 100 survival related genes²³. These and the wound response and IGS signature were evaluated.

Contingency tables were evaluated with the p-value calculated from Chi-square test and the Cramer's V statistic. The Cramer's V statistic (value can range from 0 to 1) measures the strength of association between the two variables analyzed in the contingency table, with 1 indicating perfect association and 0 indicating no association. Values between 0.36 and 0.49 indicate a substantial relation between the signatures and values >0.50 indicate a strong relation¹².

Supplementary tables

Table S2.1 Percentages of genes in the different cell cycle phases in the two proliferation signatures (numbers are given between brackets).

| Cell cycle phase | Signature 1 | Signature 2 |
|----------------------------|-------------|-------------|
| G1/S | 18.0 (147) | 18.8 (29) |
| S | 18.0 (147) | 19.5 (30) |
| G2 | 19.5 (159) | 26.0 (40) |
| G2/M | 21.7 (177) | 22.1 (34) |
| M/G1 | 16.0 (130) | 6.5 (10) |
| Matching combinations* | 3.1 (25) | 3.9 (6) |
| Non-matching combinations† | 3.7 (30) | 3.3 (5) |

* Different cloneIDs for 1 UnigenelD are found in different phases, but all phases represent the same proliferation status (i.e S, G₂, G₂/M)

† Different cloneIDs for 1 UnigenelD are found in different phases and the phases represent a different proliferation status (i.e G₂/M, G₁/S)

Table S2.2 Results of log-rank test for the signatures for the different endpoints.

| Signature 1 | | | | | |
|-----------------|---------------|----------------------|-----------------------|---|---|
| Dataset | Endpoint | P-value | SD | Range | % significant runs |
| Miller | 5 years | 3.1×10^{-3} | 1.6×10^{-3} | $1.5 \times 10^{-3} - 4.7 \times 10^{-3}$ | 100 |
| | 10 years | 6.9×10^{-4} | 3.4×10^{-4} | $3.6 \times 10^{-4} - 1.0 \times 10^{-3}$ | 100 |
| Wang | 5 years | 1.9×10^{-3} | 1.0×10^{-3} | $9.9 \times 10^{-4} - 3.0 \times 10^{-3}$ | 100 |
| | van de Vijver | 5 years | 4.6×10^{-5} | 0.0 | $4.6 \times 10^{-5} - 4.6 \times 10^{-5}$ |
| Zhao | 5 years | 2.5×10^{-7} | 0.0 | $2.5 \times 10^{-7} - 2.5 \times 10^{-7}$ | 100 |
| | 10 years | 0.48 | 8.0×10^{-2} | 0.39 - 0.55 | 0 |
| Beer | 10 years | 0.54 | 3.0×10^{-2} | 0.51 - 0.57 | 0 |
| | 2 years | 0.16 | 5.4×10^{-2} | 0.12 - 0.20 | 0 |
| | 5 years | 0.49 | 9.8×10^{-2} | 0.11 - 0.63 | 0 |
| Signature 2 | | | | | |
| Dataset | Endpoint | P-value | SD | Range | % significant runs |
| Miller | 5 years | 4.1×10^{-3} | 1.8×10^{-3} | $2.0 \times 10^{-3} - 6.2 \times 10^{-3}$ | 100 |
| | 10 years | 7.0×10^{-4} | 1.8×10^{-4} | $4.6 \times 10^{-4} - 9.3 \times 10^{-4}$ | 100 |
| Wang | 5 years | 2.3×10^{-3} | 1.8×10^{-4} | $5.7 \times 10^{-5} - 6.4 \times 10^{-4}$ | 100 |
| | van de Vijver | 5 years | 1.4×10^{-6} | 0.0 | $1.4 \times 10^{-6} - 1.4 \times 10^{-6}$ |
| Zhao | 10 years | 3.0×10^{-8} | 1.6×10^{-10} | $3.0 \times 10^{-8} - 3.0 \times 10^{-8}$ | 100 |
| | 5 years | 3.1×10^{-2} | 1.1×10^{-2} | $1.9 \times 10^{-2} - 4.2 \times 10^{-2}$ | 100 |
| Beer | 10 years | 2.3×10^{-2} | 3.2×10^{-3} | $2.0 \times 10^{-2} - 2.7 \times 10^{-2}$ | 100 |
| | 2 years | 3.3×10^{-3} | 2.2×10^{-5} | $3.3 \times 10^{-3} - 3.4 \times 10^{-3}$ | 100 |
| | 5 years | 2.8×10^{-2} | 6.8×10^{-5} | $2.8 \times 10^{-2} - 2.8 \times 10^{-2}$ | 100 |
| Wound signature | | | | | |
| Dataset | Endpoint | P-value | SD | Range | % significant runs |
| Miller | 5 years | 1.8×10^{-4} | 1.1×10^{-4} | $4.8 \times 10^{-5} - 8.7 \times 10^{-4}$ | 100 |
| | 10 years | 9.3×10^{-6} | 3.9×10^{-6} | $3.1 \times 10^{-6} - 3.9 \times 10^{-5}$ | 100 |
| Wang | 5 years | 2.1×10^{-3} | 1.4×10^{-3} | $5.1 \times 10^{-4} - 4.1 \times 10^{-3}$ | 100 |
| | van de Vijver | 5 years | 5.9×10^{-5} | 0.0 | $5.9 \times 10^{-5} - 5.9 \times 10^{-5}$ |
| Zhao | 10 years | 9.3×10^{-7} | 5.2×10^{-9} | $9.3 \times 10^{-7} - 9.4 \times 10^{-7}$ | 100 |
| | 5 years | 2.1×10^{-2} | 6.9×10^{-3} | $2.5 \times 10^{-2} - 3.1 \times 10^{-2}$ | 100 |
| Beer | 10 years | 2.8×10^{-2} | 2.7×10^{-3} | $1.4 \times 10^{-2} - 2.8 \times 10^{-2}$ | 100 |
| | 2 years | 6.9×10^{-2} | 5.1×10^{-2} | $1.5 \times 10^{-2} - 0.10$ | 49 |
| | 5 years | 0.12 | 1.2×10^{-2} | $8.1 \times 10^{-2} - 0.13$ | 0 |
| IGS signature | | | | | |
| Dataset | Endpoint | P-value | SD | Range | % significant runs |
| Miller | 5 years | 5.3×10^{-3} | 2.6×10^{-3} | $2.8 \times 10^{-3} - 8.1 \times 10^{-3}$ | 100 |
| | 10 years | 2.4×10^{-3} | 7.4×10^{-4} | $1.7 \times 10^{-3} - 3.2 \times 10^{-3}$ | 100 |
| Wang | 5 years | 2.1×10^{-6} | 1.7×10^{-6} | $4.3 \times 10^{-7} - 5.1 \times 10^{-6}$ | 100 |
| | van de Vijver | 5 years | 1.5×10^{-4} | 0.0 | $1.5 \times 10^{-4} - 1.5 \times 10^{-4}$ |
| Zhao | 10 years | 4.9×10^{-7} | 0.0 | $4.9 \times 10^{-7} - 4.9 \times 10^{-7}$ | 100 |
| | 5 years | 2.2×10^{-3} | 1.0×10^{-3} | $1.0 \times 10^{-3} - 3.4 \times 10^{-3}$ | 100 |
| Beer | 10 years | 4.7×10^{-3} | 9.5×10^{-4} | $3.3 \times 10^{-3} - 5.9 \times 10^{-3}$ | 100 |
| | 2 years | 2.8×10^{-2} | 6.6×10^{-3} | $2.3 \times 10^{-2} - 3.7 \times 10^{-2}$ | 100 |
| | 5 years | 5.4×10^{-2} | 5.4×10^{-3} | $5.0 \times 10^{-2} - 6.1 \times 10^{-2}$ | 61 |

Table S2.3 AUCs of individual clinical parameters and proliferation signature 2.

| Miller | Wang | | van de Vijver | | |
|--------------------|-------------|-----------------|----------------------|-----------------|------|
| Age | 0.48 | ER-status | 0.62 | Age | 0.42 |
| Elston grade | 0.65 | Proliferation** | 0.59 | Elston grade | 0.72 |
| Tumor size* | 0.71 | | | Tumor size* | 0.67 |
| ER-status | 0.58 | | | ER-status | 0.41 |
| LNS‡ | 0.70 | | | LNS‡ | 0.52 |
| PgR§ | 0.54 | | | NIH risk | 0.66 |
| P53-status | 0.57 | | | Mastectomy | 0.57 |
| Proliferation** | 0.67 | | | Chemotherapy | 0.52 |
| | | | | Hormone therapy | 0.53 |
| | | | | Proliferation* | 0.72 |
| Zhao | Beer | | | | |
| Age | 0.43 | Age | 0.60 | | |
| Sex | 0.45 | Sex | 0.45 | | |
| Performance status | 0.60 | Smoking† | 0.49 | | |
| Grade | 0.64 | Stage | 0.66 | | |
| Stage | 0.85 | Differentiation | 0.59 | | |
| Proliferation** | 0.56 | K-ras mutation | 0.52 | | |
| | | Proliferation** | 0.64 | | |

* Categories: ≤ 2 cm or > 2 cm

† Categories: smoker or non-smoker

‡ LNS: lymph-node status

§ PgR: progesterone receptor status

** proliferation: proliferation signature 2

Table S2.4 Parameters selected with stepwise backward selection in multivariate Cox-regression analyses including signature 2, wound and IGS signature.

| Input | | | | | | |
|---------------------------|----------|------------------------------------|-----------------------------|---------------------------------------|------------------------------------|---------------------|
| Dataset | Endpoint | Clinical Parameters | Clinical parameters + Prol+ | Clinical parameters + Wound signature | Clinical parameters + Prol+ + IGS | Signatures |
| Miller | 5 years | Tumor size* LNS^ | Tumor size* Prol+ | Tumor size* Age LNS^ | Tumor size* Age LNS^ | Age Wound Prol+ |
| | 10 years | Tumor size* LNS^ P53-status | Tumor size* Prol+ | Tumor size* Age LNS^ | Tumor size* Age LNS^ | Prol+ Wound IGS |
| Wang van de Vijver | 5 years | ER-status | Prol+ | Prol+ | IGS | IGS |
| | 5 years | Tumor size* ER-status Elston grade | Tumor size* Elston grade | Tumor size* ER-status Elston grade | Tumor size* ER-status Elston grade | ER-status Prol+ IGS |
| Zhao | 10 years | Tumor size* ER-status Elston grade | Tumor size* Elston grade | Tumor size* Age Elston grade Prol+ | Tumor size* Age Elston grade | Age IGS Prol+ IGS |
| | 5 years | Stage Grade Perform stat§ | Stage Perform stat§ | Stage Perform | Stage Perform stat§ | Prol+ IGS Wound IGS |
| Beer | 10 years | Stage Perform stat§ | Stage Perform stat§ | Stage Perform | Stage Perform stat§ | Prol+ Wound IGS |
| | 2 years | Stage Diff | Stage | Stage Diff | Stage | Diff Prol+ |
| | 5 years | Stage Diff Age | Stage Age | Stage Pro+ Age | Stage Age | Prol+ Pro+ |

* Categories: ≤ 2 cm or > 2 cm
 ^ LNS: lymph-node status
 + prol: proliferation signature 2
 § Perf stat: performance status

Chapter 2

Table S2.5 Two-way contingency table analyses measuring the association between the proliferation signature and the wound (A), IGS (B) and 32-gene p53 signature (C) in the Miller dataset.

| A | | Wound signature | |
|--|------------|-----------------|-----------|
| | | Quiescent | Activated |
| Proliferation signature | Low score | 120 | 15 |
| | High score | 19 | 82 |
| Statistics: Cramer's V = 0.70, p-value < 0.001 | | | |

| B | | IGS signature | |
|--|------------|---------------|------------|
| | | Low score | High score |
| Proliferation signature | Low score | 105 | 30 |
| | High score | 19 | 82 |
| Statistics: Cramer's V = 0.58, p-value < 0.001 | | | |

| C | | 32-gene p53 signature | |
|--|------------|-----------------------|-------------|
| | | p53-wt like | P53-mu like |
| Proliferation signature | Low score | 125 | 43 |
| | High score | 10 | 58 |
| Statistics: Cramer's V = 0.55, p-value < 0.001 | | | |

Table S2.6 Two-way contingency table analyses measuring the association between the proliferation signature and the wound (A) and IGS (B) signature in the Wang dataset.

| A | | Wound signature | |
|---|------------|-----------------|-----------|
| | | Quiescent | Activated |
| Proliferation signature | Low score | 124 | 12 |
| | High score | 60 | 90 |
| Statistics: Cramer's V = 0.53 p-value < 0.001 | | | |

| B | | IGS signature | |
|--|------------|---------------|------------|
| | | Low score | High score |
| Proliferation signature | Low score | 91 | 45 |
| | High score | 27 | 123 |
| Statistics: Cramer's V = 0.50, p-value < 0.001 | | | |

Prognostic value of a proliferation signature

Table S2.7 Two-way contingency table analyses measuring the association between the proliferation signature and the wound (A), IGS (B) and 70-gene (C) signature in the van de Vijver dataset.

| A | | Wound signature | |
|--|------------|-----------------|-----------|
| | | Quiescent | Activated |
| Proliferation signature | Low score | 150 | 10 |
| | High score | 25 | 110 |
| Statistics: Cramer's V = 0.76, p-value < 0.001 | | | |

| B | | IGS signature | |
|--|------------|---------------|------------|
| | | Low score | High score |
| Proliferation signature | Low score | 114 | 46 |
| | High score | 34 | 101 |
| Statistics: Cramer's V = 0.46, p-value < 0.001 | | | |

| C | | 70-gene signature | |
|--|------------|-------------------|--------------|
| | | Poor profile | Good profile |
| Proliferation signature | Low score | 106 | 54 |
| | High score | 9 | 126 |
| Statistics: Cramer's V = 0.61, p-value < 0.001 | | | |

Table S2.8 Two-way contingency table analyses measuring the association between the proliferation signature and the wound (A) and IGS (B) signature in the Zhao dataset.

| A | | Wound signature | |
|--|------------|-----------------|-----------|
| | | Quiescent | Activated |
| Proliferation signature | Low score | 84 | 11 |
| | High score | 49 | 33 |
| Statistics: Cramer's V = 0.33, p-value < 0.001 | | | |

| B | | IGS signature | |
|--|------------|---------------|------------|
| | | Low score | High score |
| Proliferation signature | Low score | 77 | 18 |
| | High score | 30 | 52 |
| Statistics: Cramer's V = 0.45, p-value < 0.001 | | | |

Table S2.9 Two-way contingency table analyses measuring the association between the proliferation signature and the wound (A), IGS (B) and 100 survival related genes (C) in the Beer dataset.

| A | | Wound signature | |
|--|------------|----------------------------|------------|
| | | Quiescent | Activated |
| Proliferation signature | Low score | 42 | 3 |
| | High score | 6 | 35 |
| Statistics: Cramer's V = 0.78, p-value < 0.001 | | | |
| B | | IGS signature | |
| | | Low score | High score |
| Proliferation signature | Low score | 39 | 6 |
| | High score | 14 | 27 |
| Statistics: Cramer's V = 0.54, p-value < 0.001 | | | |
| C | | 100 survival related genes | |
| | | Low risk | High risk |
| Proliferation signature | Low score | 33 | 12 |
| | High score | 10 | 31 |
| Statistics: Cramer's V = 0.50, p-value < 0.001 | | | |

References

1. Bild, A.H., Potti, A. & Nevins, J.R. Linking oncogenic pathways with therapeutic opportunities. *Nat Rev Cancer* 6, 735-741 (2006).
2. Chen, H.Y., et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 356, 11-20 (2007).
3. Dai, H., et al. A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. *Cancer Res* 65, 4059-4066 (2005).
4. Quackenbush, J. Microarray analysis and tumor classification. *N Engl J Med* 354, 2463-2472 (2006).
5. Dupuy, A. & Simon, R.M. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* 99, 147-157 (2007).
6. Chang, H.Y., et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* 2, E7 (2004).
7. Chi, J.T., et al. Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. *PLoS Med* 3, e47 (2006).
8. Sung, F.L., et al. Genome-wide expression analysis using microarray identified complex signaling pathways modulated by hypoxia in nasopharyngeal carcinoma. *Cancer Lett* (2007).
9. Winter, S.C., et al. Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. *Cancer Res* 67, 3441-3449 (2007).
10. Liu, R., et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 356, 217-226 (2007).
11. Adler, A.S., et al. Genetic regulators of large-scale transcriptional signatures in cancer. *Nat Genet* 38, 421-430 (2006).
12. Fan, C., et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 355, 560-569 (2006).
13. van de Vijver, M.J., et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347, 1999-2009 (2002).
14. Bourhis, J., et al. Hyperfractionated or accelerated radiotherapy in head and neck cancer: a meta-analysis. *Lancet* 368, 843-854 (2006).
15. De Ruyscher, D., et al. Time between the first day of chemotherapy and the last day of chest radiation is the most important predictor of survival in limited-disease small-cell lung cancer. *J Clin Oncol* 24, 1057-1063 (2006).
16. Whitfield, M.L., George, L.K., Grant, G.D. & Perou, C.M. Common markers of proliferation. *Nat Rev Cancer* 6, 99-106 (2006).
17. Perou, C.M., et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci U S A* 96, 9212-9217 (1999).
18. Sotiriou, C., et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 98, 262-272 (2006).
19. Paik, S., et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351, 2817-2826 (2004).
20. van 't Veer, L.J., et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530-536 (2002).

21. Whitfield, M.L., *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* 13, 1977-2000 (2002).
22. Wang, Y., *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365, 671-679 (2005).
23. Beer, D.G., *et al.* Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 8, 816-824 (2002).
24. Miller, L.D., *et al.* An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A* 102, 13550-13555 (2005).
25. Dyrskjot, L., *et al.* A molecular signature in superficial bladder carcinoma predicts clinical outcome. *Clin Cancer Res* 11, 4029-4036 (2005).
26. Larsen, J.E., *et al.* Gene expression signature predicts recurrence in lung adenocarcinoma. *Clin Cancer Res* 13, 2946-2954 (2007).
27. Rosenwald, A., *et al.* The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer Cell* 3, 185-197 (2003).
28. Corvo, R., *et al.* Evidence of cell kinetics as predictive factor of response to radiotherapy alone or chemoradiotherapy in patients with advanced head and neck cancer. *Int J Radiat Oncol Biol Phys* 47, 57-63 (2000).
29. Gasinska, A., Fowler, J.F., Lind, B.K. & Urbanski, K. Influence of overall treatment time and radiobiological parameters on biologically effective doses in cervical cancer patients treated with radiation therapy alone. *Acta Oncol* 43, 657-666 (2004).
30. Begg, A.C., *et al.* The value of pretreatment cell kinetic parameters as predictors for radiotherapy outcome in head and neck cancer: a multicenter analysis. *Radiother Oncol* 50, 13-23 (1999).
31. Caly, M., *et al.* Analysis of correlation between mitotic index, MIB1 score and S-phase fraction as proliferation markers in invasive breast carcinoma. Methodological aspects and prognostic value in a series of 257 cases. *Anticancer Res* 24, 3283-3288 (2004).
32. Daniels, J.M., *et al.* Mitotic index does not predict prognosis in stage IA non-small cell lung cancer. *Lung Cancer* 38, 163-167 (2002).
33. Jalava, P., *et al.* Ki67 immunohistochemistry: a valuable marker in prognostication but with a risk of misclassification: proliferation subgroups formed based on Ki67 immunoreactivity and standardized mitotic index. *Histopathology* 48, 674-682 (2006).
34. Zhao, H., *et al.* Gene expression profiling predicts survival in conventional renal cell carcinoma. *PLoS Med* 3, e13 (2006).

Chapter 3

The use of a comprehensive tumour xenograft dataset to validate gene signatures relevant for radiation response

Published in: Radiotherapy and Oncology 2009; 92(3): 417–422

Maud H.W. Starmans, Daniel Zips, Bradly G. Wouters, Michael Baumann, Philippe Lambin

Abstract

Purpose: To investigate the use of xenograft models in a novel gene signature validation method using gene expression microarrays.

Material and methods: Gene expression profiles of ten human Head and Neck squamous cell carcinomas (HNSCCs) were obtained. Several published prognostic gene expression signatures were evaluated within this set. These consisted of different radiotherapy relevant signatures (i.e. for hypoxia, proliferation and 'stemness'). Signatures were correlated with various endpoints that have been determined in the ten different xenograft models. These include immunohistochemical measures for hypoxia and proliferation, volume doubling time (VDT) and local tumour control after fractionated irradiation or after single dose irradiation under clamp hypoxia.

Results: We found several significant correlations between the published gene expression signatures and tumour parameters. Several signatures, like the proliferation and wound signature correlated with BrdU labelling index. Further a 'stemness' related gene signature showed a strong negative correlation with hypoxic fraction.

Conclusions: Simultaneous assessment of immunohistochemistry, in vivo tumour properties and gene expression profiling in a comprehensive set of xenograft models can be used to validate and potentially infer biological information about prognostic gene signatures.

Introduction

Within oncology, we are challenged with the fact that tumours with similar characteristics (i.e. stage, grade and size) show striking differences in response to the same therapy. Understanding the underlying causes of these differences is crucial if we want to improve and individualize cancer treatment. Over the last few years, expression microarrays have been used to identify sets of genes, or so-called signatures that can improve diagnosis and risk stratification^{1,2}. Often these signatures are derived directly from clinical samples using supervised analysis methods to create gene sets that can identify patient groups with distinct risk properties¹. One of the drawbacks of this method is that these signatures are often difficult to interpret with respect to the underlying biological mechanisms that may be responsible for this categorization^{3,4}. Conversion of these prognostic signatures into predictive tests is therefore not straightforward. An alternative approach utilizes *in vitro* expression profiling to identify gene sets that are responsive to a specific biological process. Several studies have identified signatures with this procedure which are associated with prognosis⁵⁻⁸.

We hypothesize that a signature evaluation step in an animal model *in vivo* might provide novel insight for both of these methodologies. A signature can be evaluated in a set of tumour xenograft models, in which in parallel other relevant tumour features are assessed. Immunohistochemistry can be implemented to assess for example hypoxic fraction (HF) and proliferation. Furthermore, volume doubling time (VDT) and the radiation dose necessary to locally control 50% of the tumours (TCD₅₀) are radiotherapy treatment relevant variables. Comparison of various signatures with these phenotypic parameters can potentially provide unique information into the underlying biological mechanisms that confer prognostic power to the signatures. This is especially true for signatures derived from clinical studies for which no *a priori* expectation of its biological basis is available. For an *in vitro* derived signature this can be used as an extra *in vivo* validation step.

Proliferation, hypoxia and 'stemness' are parameters implicated in determining outcome of radiotherapy⁹⁻¹⁴. Gene signatures related to these phenotypes may thus provide valuable information on prognosis and prediction for patients treated with radiotherapy. Here we have explored the possibility of validating several of these published signatures related to these features on a comprehensive dataset of human tumour xenografts.

Materials and methods

Animals and tumour models

The experiments were performed using 7 to 14 week-old male and female NMRI-*nu* (*nu/nu*) mice (Experimental Centre of the Medical Faculty of the University of Dresden). Animal facilities and experiments were approved according to the German animal welfare regulations. Ten human head and neck squamous cell carcinomas (HNSCC) were used. These have been described in detail previously see Table 3.1^{12,15-17}. For the experiments, source tumours were cut into small pieces and transplanted subcutaneously into the right hind-leg of anaesthetized mice (120 mg/kg body weight (b.w.) ketamine intraperitoneal (i.p.) and 16 mg/kg b.w. xylazine i.p.). Volume doubling time (VDT) were calculated as previously described, data on FaDU, UT-SCC-5, UT-SCC-14, UT-SCC-15 and XF354 were previously reported¹⁵.

Tumour irradiation and analysis of local control data

Irradiation experiments have been previously reported in detail^{12,15,16}. Briefly, local irradiations were given with 200 kV X-rays (0.5 mm Cu filter, 1 Gy/min) when the tumour reached a diameter of 7mm (range: 7–9mm). Fractionated irradiations were given under ambient conditions with 30 doses within 6 weeks. For single dose irradiation under homogeneous hypoxia a heavy clamp was placed over the proximal thigh of anaesthetized mice 2 min before and during irradiation. For each xenograft model TCD₅₀ was calculated for both schemes^{12,15,16}.

Immunohistochemistry

Histological evaluation of hypoxia¹⁵⁻¹⁷ and tumour cell proliferation (Zips *et al*, in press¹⁸) have been previously reported in detail. Briefly, tumour bearing animals were injected with the hypoxia cell marker pimonidazole (Natural Pharmacia International Inc., Research triangle park, NC) and the S-phase marker BrdU (Serva, Heidelberg, Germany), both 0.1 mg/g b.w. i.p. dissolved in 10 mg/ml NaCl 0.9%, 1 hour and 15 minutes before tumour excision, respectively. For hypoxia staining, sections were incubated with rabbit anti-pimonidazole polyclonal antibody (kindly provided by Dr. J. Raleigh, Institute of Radiation Oncology, University of North Carolina Medical School, Chapel Hill, NC) followed by a secondary goat anti-rabbit FITC polyclonal antibody (Jackson ImmunoResearch, West Grove, PA). The hypoxic fraction (HF) was determined as the pimonidazole positive area per viable tumour

area. For detection of proliferation, sections were probed with biotinylated monoclonal anti-BrdU antibody (DakoCytomation, Glostrup, Denmark) followed by addition of streptavidin-peroxidase complex and DAB chromogen. BrdU LI was determined by manually counting positive and negative cells per high-power visual field (400-fold) in 10 fields. Five tumours (2 central cross-sections per tumour) were evaluated per tumour model. One tumour for CAL33 did not show positive cells and was left out. The staining procedure failed in HSC-4 and UT-SCC-45 tumours.

Table 3.1 Summary of previously determined tumour parameters

| Tumour type | Immunohistochemistry | TCD ₅₀ ([Gy]) (95% CI) | |
|-------------------------|--|-----------------------------------|-------------------|
| | Hypoxic fraction (HF [%]) mean ± sd | Fractionated radiation | Single dose clamp |
| CAL33 | 29.3 ± 12.8 | 105.2 (90-141) | 66.1 (54-105) |
| FaDu | 11.7 ± 5.4 | 85.2 (77-96) | 48.3 (44-53) |
| HSC4⁺ | 20.2 ± 7.1 | 82.9 (69-107) | 52.4 (46-63) |
| SAS | 21.0 ± 10.7 | 126.7 (114-140) | n.d. |
| UT-SCC-5 | 35.2 ± 8.0 | 117.2 (103-140) | 62.7 (54-74) |
| UT-SCC-8 | 14.2 ± 5.8 | 52.1 (40-61) | 44.9 (41-50) |
| UT-SCC-14 | 6.2 ± 2.1 | 52.1 (46-59) | 42.0 (37-47) |
| UT-SCC-15 | 23.6 ± 11.6 | 78.1 (63-109) | 54.1 (46-70) |
| UT-SCC-45 | 18.1 ± 7.4 | 45.4 (38-52) | 45.5 (40-55) |
| XF354 | 19.4 ± 8.2 | 47.4 (40-55) | 46.6 (39-62) |

sd: standard deviation

95% CI: 95% confidence interval

n.d.: not determined

⁺ tumour evokes residual immune response in nude mice

Data have been reported in^{12,15-17}. The HF value for UT-SCC-15 was erroneously reported in these publications.

RNA isolation, Microarrays and Signature evaluation

Total RNA was extracted with Trizol reagent (Sigma, Saint Louis, MO) according to the manufacturer's instructions. RNA quantity and quality was determined using the ND-1000 spectrometer (NanoDrop Technologies) and RNA Nano LabChip kit on the 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA), respectively.

Affymetrix HG-U133 Plus 2.0 GeneChips were used for gene expression profiling of the ten xenograft lines. Total RNA was pooled in equal amounts from 2 or 3 independent xenografts, dependent on the quality of the RNA. Biotin-labeled cRNA was synthesized from 2 µg of pooled RNA with the Affymetrix One-Cycle target

labeling. For hybridization 20 µg cRNA was used. The manufacturer's standard protocols were strictly followed (Affymetrix, Santa Clara, CA).

The Bioconductor¹⁹ packages *affy*²⁰, *affyPLM*²¹ and *simpleaffy*²¹ were used to evaluate the quality of the hybridized arrays and the quality of the pre-processing. GCRMA was performed used for background correction and normalization of the gene expression microarrays. An expression filter was applied to remove probes with low expression across the samples. With the Bioconductor package *genefilter*²¹, probes with an expression value below 100 in 50% of the arrays were excluded.

The genes in the signature were mapped to the microarray and expression data of these genes was extracted. Next, a signature score was calculated for each xenograft model. This score was defined as the weighted average expression value of the genes in the signature (Equation 1). Details on the methodology and weight assignment were previously described⁶.

$$\text{Score} = \frac{\sum_{i=1}^N w_i \cdot \text{exp}_i}{N} \quad \text{Equation (1)}$$

With: Score = signature score
 N = number of genes in the signature
 i = gene i
 w_i = weight of gene i
 exp_i = gene expression of gene i

The signature score for each signature of interest was correlated with various endpoints which have been determined in the ten different xenograft models^{12,15,17}. Pearson's correlation and linear regression were performed (p < 0.05 was considered significantly different).

Results

A comprehensive HNSCCs xenograft set was used to introduce a novel gene signature evaluation procedure. Gene expression microarray profiling was performed for xenografts grown from 10 HNSCC lines. Per model three individual tumours were grown. RNA was isolated and pooled in equal amounts for each tumour type. For CAL33, UT-SCC-14 and UT-SCC-45 RNA was pooled from two tumours, because RNA quality was not sufficient in the third tumour.

Table 3.2 Signature overview

| Gene signature | Reference | Number of genes | Phenotype |
|------------------|--------------------------------------|-----------------|----------------|
| Chi hypoxia | Chi <i>et al</i> ⁷ | 171 | Hypoxia |
| 0% early hypoxia | Seigneuric <i>et al</i> ⁸ | 79 | Hypoxia |
| 2% early hypoxia | Seigneuric <i>et al</i> ⁸ | 36 | Hypoxia |
| Wound signature | Chang <i>et al</i> ⁵ | 454 | Wound response |
| IGS | Liu <i>et al</i> ¹³ | 176 | Stemness |
| Proliferation | Starmans <i>et al</i> ⁶ | 110 | Proliferation |
| Chung signature | Chung <i>et al</i> ²² | 42 | - |

Several published gene expression signatures were evaluated in this xenograft dataset. Since proliferation, hypoxia and ‘stemness’ are implicated in determining outcome of radiotherapy, signatures related to these phenotypes were included in this study. We also included a prognostic signature specific for surgery and chemo-radiation treated Head and Neck cancer patients^{14,22}. An overview of the studied signatures is given in Table 3.2. For each xenograft model a signature score was calculated for each signature (Equation 1). This signature score was then correlated to different phenotypic tumour features.

Table 3.3 Summary of determined tumour parameters

| Tumour type | Immunohistochemistry | Tumour growth rate |
|-------------------|---|---|
| | BrdU labelling index (BrdU LI [%]) mean \pm sd | Volume doubling time (VDT [days]) median (10-90 percentiles) |
| CAL33 | 9.56 \pm 1.9 | 4.26 (3-6) |
| FaDu | 11.88 \pm 5.8 | 4.48 (3-6) |
| HSC4 ⁺ | n.d. | 3.80 (2-6) |
| SAS | 22.57 \pm 6.9 | 4.15 (3-6) |
| UT-SCC-5 | 11.63 \pm 3.7 | 6.79 (5-9) |
| UT-SCC-8 | 9.23 \pm 3.0 | 6.23 (4-9) |
| UT-SCC-14 | 6.48 \pm 1.8 | 4.18 (3-7) |
| UT-SCC-15 | 11.21 \pm 2.8 | 6.03 (4-8) |
| UT-SCC-45 | n.d. | 6.91 (4-10) |
| XF354 | 10.62 \pm 3.0 | 8.69 (5-12) |

sd = standard deviation

n.d. = not determined

⁺ tumour evokes residual immune response in nude mice

Several tumour features were previously studied for this set of xenografts. HF was earlier determined by immunohistochemistry with pimonidazole^{15,17}. Further Yaromina *et al*¹² previously studied the TCD₅₀ after fractionated irradiation (30

fractions, 6 weeks) or under clamping conditions (single dose). These parameters are displayed in Table 3.1. For this study two additional parameters were calculated. VDT and BrdU LI were assessed for each xenograft line. Data of these measurements are given in Table 3.3.

We first tested whether proliferation signatures accurately reflect more direct measures of tumour growth such as BrdU LI and VDT. We postulated that both the proliferation and Wound signature would correlate with BrdU LI. For the proliferation signature this does not need further explanation; however for the Wound signature this hypothesis is not immediately obvious. We have previously shown that the proliferation and Wound signature identify the same patients in several large microarray patient datasets⁶, which could indicate that these signatures are measures for the same process, in this case proliferation. Figure 3.1A and 3.1B show the association between BrdU LI and the proliferation and the Wound signature, respectively. Both signatures show a strong correlation with BrdU LI (r : 0.66 and 0.72, respectively). For the Wound signature this correlation is significant ($p = 0.044$). The proliferation signature only shows a borderline significant result ($p = 0.077$). For both signatures no correlation was found with VDT (proliferation signature: r : 0.22, $p = 0.550$, Wound signature: r : -0.12, $p = 0.739$). For completeness correlation coefficients were also calculated for the other tumour parameters, no additional correlations were found (Table 3.4).

Table 3.4 Results of correlating gene signatures to xenograft features

| Gene signature | Immunohistochemistry | | | | In vivo properties | | | | | |
|-------------------------|----------------------|-------|---------|-------|--------------------|-------|-------|-------|------------|-------|
| | HF | | BrdU LI | | VDT | | TCD50 | | TCD50clamp | |
| | r | p | r | p | r | p | r | p | r | p |
| Chi hypoxia | 0.01 | 0.973 | -0.47 | 0.245 | -0.35 | 0.324 | 0.08 | 0.825 | 0.22 | 0.579 |
| 0% early hypoxia | 0.04 | 0.903 | 0.87 | 0.005 | -0.19 | 0.604 | 0.45 | 0.194 | 0.01 | 0.988 |
| 2% early hypoxia | -0.18 | 0.625 | 0.62 | 0.103 | 0.23 | 0.522 | 0.05 | 0.896 | -0.34 | 0.378 |
| Wound signature | -0.18 | 0.628 | 0.72 | 0.044 | -0.12 | 0.739 | 0.21 | 0.558 | -0.26 | 0.504 |
| IGS | -0.64 | 0.048 | 0.35 | 0.394 | -0.36 | 0.304 | -0.05 | 0.884 | -0.51 | 0.161 |
| Proliferation | -0.02 | 0.959 | 0.66 | 0.077 | 0.22 | 0.550 | 0.08 | 0.825 | -0.19 | 0.620 |
| Chung signature | 0.33 | 0.360 | 0.30 | 0.477 | 0.04 | 0.921 | 0.23 | 0.521 | 0.25 | 0.510 |

r : Pearson's coefficient of correlation

p : p-value for linear regression

Further we hypothesized that the different hypoxia signatures would correlate with HF. Pearson's coefficients of correlation indicated that there was no association between the three hypoxia signatures and HF (Table 3.4, Figure 3.2A-C).

Correlations between the signatures and all other tumour features were calculated to investigate whether the hypoxia signatures were associated with other tumour parameters. The 0% early hypoxia signature showed a strong significant correlation with BrdU LI (Figure 3.1C, $r: 0.87$, $p = 0.005$). Also the 2% early hypoxia signature related to BrdU LI (Figure 3.1D, $r: 0.62$, $p = 0.103$). The Chi hypoxia signature, unlike the other two hypoxia signatures showed a negative association with BrdU LI (Figure 3.1E, $r: -0.47$, $p = 0.245$).

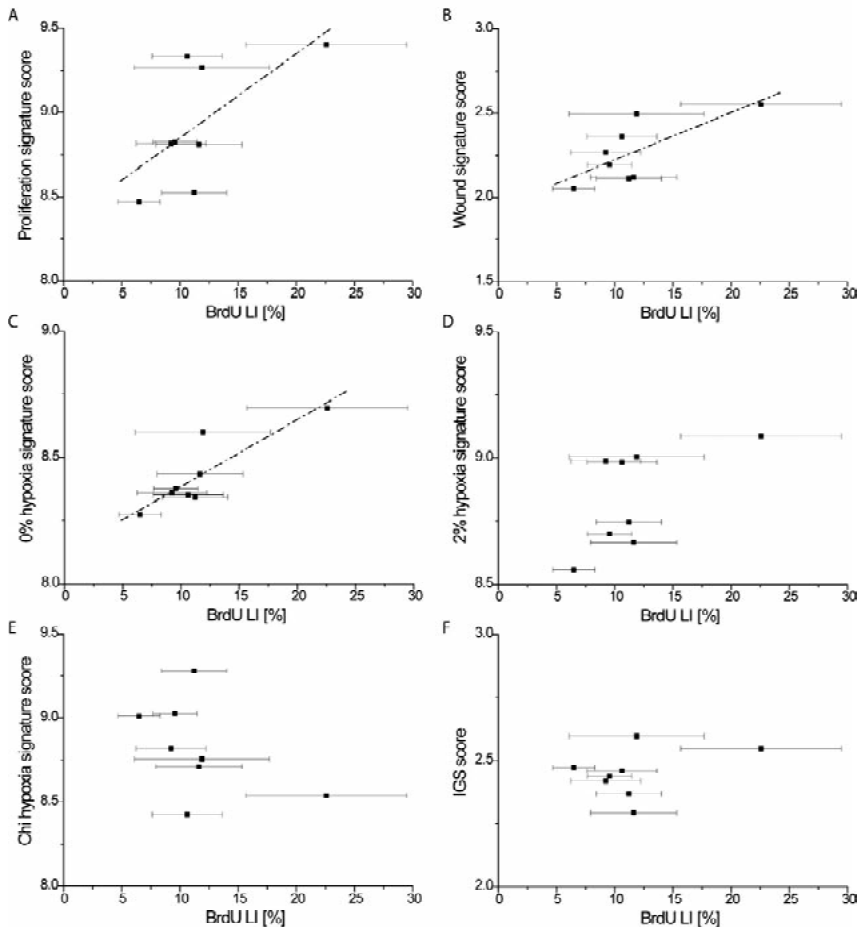


Figure 3.1

Correlation between BrdU LI and the proliferation signature (A), the Wound signature (B), the 0% early hypoxia signature (C), the 2% early hypoxia signature (D), the Chi hypoxia signature (E) and the IGS signature (F) respectively. Dotted lines indicate linear fit for (border) significant correlations.

Further for the 0% hypoxia signature a trend with TCD_{50} (fractionated irradiation) was observed, although this was not significant ($r: 0.45, p = 0.194$). The other hypoxia signatures showed no correlation at all with TCD_{50} (Table 3.4).

For the invasiveness gene signature (IGS), a ‘stemness’ related signature, and Chung signature no hypotheses were phrased. We only speculated that the IGS signature would not correlate with BrdU LI. Our group previously demonstrated in a set of large patient microarray datasets that the IGS and proliferation signature identify different patients⁶. This could indicate that these signatures track different processes. To explore what features might be associated with the IGS and Chung signature, correlations were calculated between the signatures and all tumour phenotype parameters (Table 3.4). No substantial relation between the IGS signature and BrdU LI was found (Figure 3.1F, $r: 0.35, p = 0.394$). Results showed a significant negative correlation between the IGS signature and HF (Figure 3.2D, $r: -0.64, p = 0.048$) and a negative association with TCD_{50} under clamping conditions ($r: -0.51, p = 0.161$). No correlations were found for the Chung signature (Table 3.4).

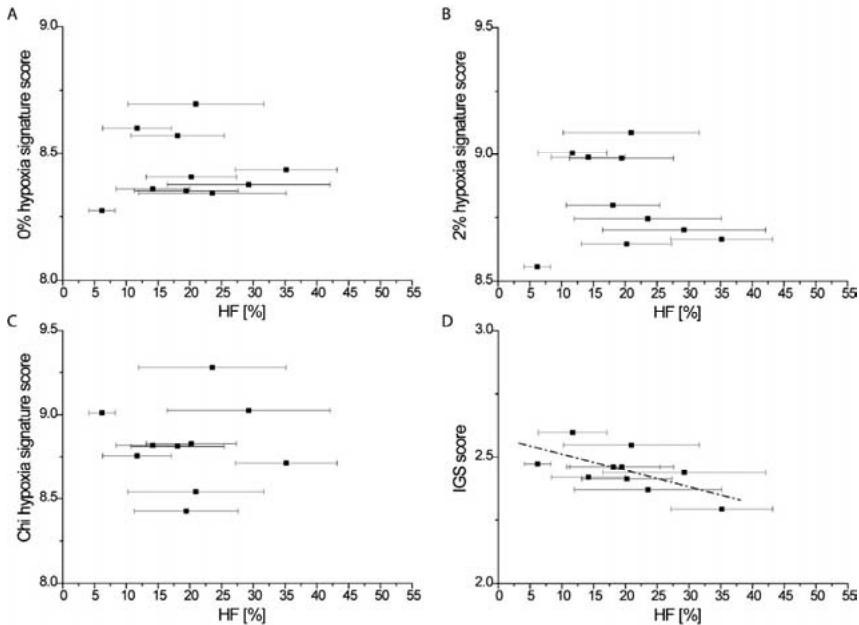


Figure 3.2

Correlation between HF and the 0% early hypoxia signature (A), the 2% early hypoxia signature (B), the Chi hypoxia signature (C) and the IGS signature (D) respectively. Dotted lines indicate linear fit for (border) significant correlations.

Discussion

Better understanding of tumour behaviour is necessary to improve and individualize cancer treatment. To date, the most important factors used to predict outcome and choose therapy modalities are tumour size, site, stage and grade. However, these factors cannot explain biologic behaviour nor treatment response.

Published gene signatures were evaluated in a comprehensive human tumour xenograft dataset, consisting of 10 different HNSCC cell lines. We included gene signatures related to proliferation, hypoxia and 'stemness', since these phenotypes are important for outcome of radiotherapy⁹⁻¹⁴. We show that both the proliferation signature and the Wound signature correlate with BrdU LI. It was previously suggested that many signatures track a common set of biologic phenotypes²³. Further Whitfield *et al*²⁴ postulated that for many signatures this would be proliferation. We previously provided data that supports this idea⁶. Surprisingly two hypoxia signatures showed no association with HF, however they also showed a strong positive correlation with BrdU LI. Previously gene ontology analyses already showed that genes involved in proliferation and cell cycle were enriched in these signatures⁸. This suggests that these two hypoxic signatures might be rather surrogate markers for proliferation than hypoxia signatures. However the IGS signature, a 'stemness' related signature, showed a strong negative correlation with HF. This appears contradictory to recent suggestions that hypoxia may confer stem-like properties or provide a niche for stem cells^{25,26}. This result will need to be further investigated.

Here gene expression profiling, immunohistochemistry and assessment of *in vivo* tumour properties were performed in different xenografts within one experiment. We tried to account for this by pooling RNA from three individual tumours for microarray profiling. Ideally, all measurements should be obtained from the same set of tumours. Due to this and the small number of microarrays used, probably only strong correlations can be identified. Signatures associated with multiple phenotypes then only demonstrate weak associations. This might explain why no correlations were found for the Chung signature.

We introduce a new signature evaluation step that could be applied to overcome part in the shortcomings of gene signature research. Any signature of interest can be evaluated with this *in vivo* procedure. A signature can be compared to different immunohistochemical parameters and *in vivo* xenograft properties. This will help

with the biological interpretation of signatures derived from clinical studies. For *in vitro* derived signatures it introduces an extra validation step, ensuring that the signature also tracks the process of interest in a complex tumour microenvironment.

Conclusion

We show that gene expression profiling in parallel with immunohistochemistry and assessment of tumour properties in a set of xenograft models can advance clinical interpretation of gene signatures.

Acknowledgements

We acknowledge financial support from Siemens (M.H.W.S), the Dutch Cancer Society (KWF grant UM 2008-4068 to B.G.W), the CTMM framework (AIRFORCE project to M.H.W.S and P.L) the EU 6th and 7th framework program (Euroxy and Metoxia program to B.G.W and P.L), the Deutsche Forschungsgemeinschaft (PAK124, Ba 1433/4, 1433/5 to M.B. and D.Z.) and the German Federal Ministry of Education and Research (BMBF 03ZIK/OncoRay to M.B. and D.Z.).

References

1. van de Vijver, M.J., et al. A gene-expression signature as a predictor of survival in breast cancer. *The New England journal of medicine* 347, 1999-2009 (2002).
2. Bild, A.H., Potti, A. & Nevins, J.R. Linking oncogenic pathways with therapeutic opportunities. *Nature reviews* 6, 735-741 (2006).
3. Quackenbush, J. Microarray analysis and tumor classification. *The New England journal of medicine* 354, 2463-2472 (2006).
4. Dai, H., et al. A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. *Cancer research* 65, 4059-4066 (2005).
5. Chang, H.Y., et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS biology* 2, E7 (2004).
6. Starmans, M.H., et al. Robust prognostic value of a knowledge-based proliferation signature across large patient microarray studies spanning different cancer types. *British journal of cancer* 99, 1884-1890 (2008).
7. Chi, J.T., et al. Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. *PLoS medicine* 3, e47 (2006).
8. Seigneuric, R., et al. Impact of supervised gene signatures of early hypoxia on patient survival. *Radiother Oncol* 83, 374-382 (2007).
9. Bussink, J., Kaanders, J.H. & van der Kogel, A.J. Microenvironmental transformations by VEGF- and EGF-receptor inhibition and potential implications for responsiveness to radiotherapy. *Radiother Oncol* 82, 10-17 (2007).
10. Bourhis, J., et al. Hyperfractionated or accelerated radiotherapy in head and neck cancer: a meta-analysis. *Lancet* 368, 843-854 (2006).
11. Baumann, M., Krause, M. & Hill, R. Exploring the role of cancer stem cells in radioresistance. *Nature reviews* 8, 545-554 (2008).
12. Yaromina, A., et al. Pre-treatment number of clonogenic cells and their radiosensitivity are major determinants of local tumour control after fractionated irradiation. *Radiother Oncol* 83, 304-310 (2007).
13. Chen, H.Y., et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 356, 11-20 (2007).
14. Pramana, J., et al. Gene expression profiling to predict outcome after chemoradiation in head and neck cancer. *International journal of radiation oncology, biology, physics* 69, 1544-1552 (2007).
15. Yaromina, A., et al. Pimonidazole labelling and response to fractionated irradiation of five human squamous cell carcinoma (hSCC) lines in nude mice: the need for a multivariate approach in biomarker studies. *Radiother Oncol* 81, 122-129 (2006).
16. Quennet, V., et al. Tumor lactate content predicts for response to fractionated irradiation of human squamous cell carcinomas in nude mice. *Radiother Oncol* 81, 130-135 (2006).
17. Bayer, C., et al. PAI-1 levels predict response to fractionated irradiation in 10 human squamous cell carcinoma lines of the head and neck. *Radiother Oncol* 86, 361-368 (2008).
18. Zips, D., et al. Triple angiokinase inhibition, tumour hypoxia and radiation response of FaDu human squamous cell carcinomas. *Radiother Oncol in press*(2009).
19. Gentleman, R.C., et al. Bioconductor: open software development for computational biology and

- bioinformatics. *Genome biology* 5, R80 (2004).
20. Gautier, L., Cope, L., Bolstad, B.M. & Irizarry, R.A. *affy*--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics (Oxford, England)* 20, 307-315 (2004).
 21. Bolstad, B.M., et al. Quality Assessment of Affymetrix GeneChip Data. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds. Gentleman, R., Huber, W., Carey, V.J., Irizarry, R.A. & Dudoit, S.) 473 (Springer, New York, 2005).
 22. Chung, C.H., et al. Gene expression profiles identify epithelial-to-mesenchymal transition and activation of nuclear factor-kappaB signaling as characteristics of a high-risk head and neck squamous cell carcinoma. *Cancer research* 66, 8210-8218 (2006).
 23. Fan, C., et al. Concordance among gene-expression-based predictors for breast cancer. *The New England journal of medicine* 355, 560-569 (2006).
 24. Whitfield, M.L., George, L.K., Grant, G.D. & Perou, C.M. Common markers of proliferation. *Nature reviews* 6, 99-106 (2006).
 25. Hill, R.P., Marie-Egyptienne, D.T. & Hedley, D.W. Cancer stem cells, hypoxia and metastasis. *Seminars in radiation oncology* 19, 106-111 (2009).
 26. Das, B., et al. Hypoxia enhances tumor stemness by increasing the invasive and tumorigenic side population fraction. *Stem cells (Dayton, Ohio)* 26, 1818-1830 (2008).

Chapter 4

Validation of a PCR-based proliferation signature *in vitro*, *ex vivo* and in patient studies

Submitted

Maud H.W. Starmans, Natasja G. Lieuwes, Paul N. Span, Syed Haider, Ludwig Dubois, Francis Nguyen, Hanneke W. van Laarhoven, Fred C.G.J. Sweep, Bradly G. Wouters, Paul C. Boutros, Philippe Lambin

Abstract

Background: Previously we demonstrated that an mRNA signature reflecting cellular proliferation had strong prognostic value. Since clinical applicability of signatures can be controversial, we sought to improve the clinical utility of our marker by validating its biological relevance, reproducibility in independent datasets, and applicability using an independent technique, quantitative polymerase-chain-reaction (qPCR).

Methods and Findings: To facilitate signature evaluation with qPCR a novel computational procedure was used to reduce the number of signature genes without significant information loss. These genes were biologically validated in different human cancer cell lines upon serum starvation and in a panel of 168 xenografts. Expression of the qPCR-based signature was dramatically decreased under starvation conditions and inversely correlated with tumor volume doubling time in xenografts.

Analyses were then extended to large patient cohorts: 1820 breast cancers and 858 non-small-cell lung cancers (NSCLC). The signature validated in breast cancer (Hazard ratio [HR] 1.63; $P < 0.001$) and NSCLC adenocarcinoma (HR 1.64; $P < 0.001$) microarray datasets.

Lastly, qPCR in a node-negative, non-adjuvantly treated breast cancer cohort ($n = 129$) showed that patients assigned to the high proliferation group had worse disease-free survival than those in the low proliferation group (HR 2.25; $P < 0.05$). This prognosticity was most pronounced in stage I tumors (HR 5.92; $P < 0.001$).

Conclusions: We have developed and validated a qPCR-based proliferation signature. This test might be used in the clinic to select (early-stage) patients for specific treatments that target proliferation.

Introduction

Over the last decade, gene expression microarray technology has had a profound impact on cancer research. The ability to analyze the expression of thousands of genes in a single experiment has been systematically used to derive prognostic and predictive markers for many cancer types¹⁻⁴. Numerous of these “signatures” show good prognostic power, but surprisingly gene-wise overlap between them has been minimal⁵⁻⁸ which increases the difficulty of introducing microarrays in clinical practice. Moreover, studies comparing data originating from different microarray platforms have reported poor inter-platform correlations^{9,10}. Nevertheless, multiple studies in breast and non-small-cell lung cancer (NSCLC) have shown that most of these signatures exhibit similar prognostic performance and identify identical patients^{7,11}. These data suggest that, although gene-wise overlap is small, the signatures track common underlying biology that determine patient outcome. Among others, Weigelt *et al.*¹² have suggested that proliferation genes drive the prognostic power of these signatures^{11,13,14}. A large meta-analysis by Wirapati *et al.*¹⁵ supports this concept.

To determine if this result could be clinically useful, we previously developed a signature based on 104 proliferation genes¹⁶. This signature was derived from two *in vitro* gene expression datasets. Genes were selected that showed a cycling pattern after synchronization in one dataset and responded to serum stimulation in the other. Our proliferation signature exhibited strong prognostic power in several large transcriptome datasets representing different cancer types¹⁶. Further, the proliferation signature and multiple other signatures identified similar patients as having good or poor prognosis¹⁶. These results substantiate the hypothesis that many published signatures act as surrogates of proliferation.

The clinical applicability of gene expression signatures remains controversial; studies seem to lack consistency and external validation is not straightforward^{8,17-21}. Many gene expression signatures were developed since the introduction of gene expression microarray technology, however so far only in breast cancer two prognostic gene profiles are tested in large prospective trials^{12,15,22,23}. The dimensionality of gene expression microarrays makes statistical analysis complex, and large numbers of samples are required for reproducible results^{8,24}. An approach to only evaluate a select number of transcripts may therefore provide an efficient alternative to high throughput expression profiling. The use of a PCR-based test to

evaluate the proliferation signature would assist in the application to a clinical setting²⁵. Furthermore, a PCR-based technique does not necessitate the availability of fresh-frozen tissue, whereas this is recommended for gene expression microarrays²⁶. Many more samples might thus be available to validate classifiers with a PCR-based technique.

Initially we examined whether it was possible to reduce the number of genes in the proliferation signature, without deteriorating its prognostic value. The original signature consisted of 104 genes, and so reducing this number would make data collection, analysis and transfer to a PCR-based approach simpler and more transparent. To further facilitate translation of this reduced proliferation signature to a PCR-platform a series of *in vitro* and *ex vivo* validation experiments were performed. We reduced the proliferation signature to 10 genes and validated it in 1820 breast cancer and 862 NSCLC patients. Lastly the reduced proliferation signature was applied to another independent, 129-patient breast cancer cohort with qPCR to demonstrate clinical utility.

Materials and methods

Gene size reduction

In our original study¹⁶ we developed a 104-gene proliferation signature and evaluated it in five different microarray datasets representing breast, lung, and renal cancers²⁷⁻³¹. To reduce the size of the signature, each gene was tested for its univariate prognostic value in each dataset. This was for disease-specific^{27,30}, metastasis-free²⁸ or overall^{29,31} survival, depending on what was reported for the dataset. Only genes that had fewer than 25% missing values were included. Expression of each gene was used as a continuous variable as input for receiver operator curve (ROC) analysis. Genes were ranked within each dataset by the area under the receiver operator curve (AUC) (Supplementary Data File S1), and then subjected to a rank-based filtering. The filtering criteria were dependent on the number of datasets that included a certain gene, and were:

- Present in 1/5 dataset: Discard the gene
- Present in 2/5 datasets: Select when ranked in top 20 for both datasets
- Present in 3/5 datasets: Select when ranked in top 20 in all three datasets
- Present in 4/5 datasets: Select when ranked in top 20 in 3 out of 4 datasets

- Present in 5/5 datasets: Select when ranked in the top 20 in 4 out of 5 datasets

qPCR

RNA was reverse-transcribed using I-script (Bio-Rad) and quantitative PCR was performed in ABI 7500 (Applied Biosystems). Gene abundance was detected using power SYBR Green I (Applied Biosystems). Primer sequences are provided in Supplementary Data File S2. Relative abundance of every gene per sample ($X_{\text{gene, sample}}$) was calculated using standard curves and normalization to 18S rRNA signal (equation 1). This was followed by median scaling per gene for each dataset³².

$$X_{\text{gene, sample}} = \frac{C_{T_{\text{gene, sample}}}}{\text{slope}_{\text{gene}}} - \frac{C_{T_{18S, \text{sample}}}}{\text{slope}_{18S}} \quad \text{Equation (1)}$$

A multi-gene signature score was subsequently calculated for the reduced proliferation signature as follows:

$$\text{Score} = \sum_{n=1}^N \text{gene}_{\text{expr}, n} \quad \text{Equation (2)}$$

In which N is the number of genes in the multi-gene marker. The parameter $\text{gene}_{\text{expr}, n}$ for a sample equals the value 1 if the sample has a level of gene n above the median for all samples in the dataset and -1 otherwise. All data analyses were performed in R (v2.12.1).

In vitro validation

To validate the involvement in proliferation of the genes in the reduced signature, serum starvation experiments were performed in five cancer cell lines (MCF7, HeLa, HT-29, U-2 OS and DU145). Cells were grown either in normal serum containing medium (10% fetal bovine serum, FBS), the control situation, or in low serum containing medium (0.1% FBS, starvation condition) for 48 hours. RNA was isolated for both conditions for three biological replicates. The multi-gene signature score (Equation 2) was calculated for each sample. Scores were then compared between normal and serum starvation conditions with a two sample two-tailed unpaired Student's t-test (R v2.12.1).

Ex vivo validation with PCR

A large set of xenografts (n = 168) was used to assess whether it is feasible to evaluate the reduced proliferation signature in tumor material. Material was isolated from xenografts grown from different cancer cell lines (HeLa, HT-29, U-87, LS 174T, HCT 116 and Hep G2) obtained from previous studies in which tumor volume doubling times (VDTs) were calculated³³⁻³⁶. The multi-gene signature score (Equation 2) was calculated and used to median dichotomize the xenograft samples. Differences in tumor VDTs were then assessed between the two groups with a two sample two-tailed unpaired Student's t-test (R v2.12.1).

Validation in independent microarray datasets

The reduced proliferation signature was further validated in independent public mRNA abundance datasets. Several breast cancer³⁷⁻⁴⁷ and NSCLC^{1,37,48-50} datasets were used to assess the prognostic power of the reduced proliferation signature. For NSCLC the datasets reported on adenocarcinoma and/or squamous cell carcinoma. Considering these are completely different disease types, separate analyses were performed per subgroup. When overall survival was provided this was used as endpoint, otherwise disease-specific survival (or the closest variant available) was used. All datasets used Affymetrix microarrays, which were normalized using the RMA algorithm⁵¹ (R packages: affy v1.26.1) combined with updated ProbeSet annotations⁵² (R packages v12.1.0: hgu95av2hsentrezgcdf, hgu133ahsentrezgcdf, hgu133bhsentrezgcdf and hgu133plus2hsentrezgcdf). Genes were matched across datasets based on Entrez Gene IDs. Median scaling and housekeeping gene normalization (to the geometric mean of ACTB, BAT1, B2M and TBP levels) was performed³². The multi-gene signature score (equation 2) was used to median dichotomize the patients in a dataset. Patients predicted as having good or poor prognosis in any of the datasets were pooled into different groups. This was done for the breast cancer and lung cancer datasets separately. Prognostic performance of the reduced proliferation signature was evaluated by Cox proportional hazard ratio modeling followed by the Wald test (R survival package v2.36-2). For breast cancer 15-year survival was used as end-point and 5-year survival for NSCLC.

Validation in independent patient cohort

The reduced proliferation signature was evaluated using qPCR in a breast cancer patient cohort of the breast tumor bank of the Radboud University Nijmegen Medical Centre (Nijmegen, The Netherlands) as described previously⁵³. Patients underwent modified radical mastectomy or a breast-conserving lumpectomy between November 1987 and December 1997. Postoperative radiotherapy was administered, to the breast after an incomplete resection or after breast-conserving treatment, or to the parasternal lymph nodes when the tumor was medially localized. Patients did not receive (neo-) adjuvant systemic therapy according to the standard treatment protocol at the time. RNA was available from 129 lymph-node negative breast cancer patients.

qPCR was carried out to evaluate the reduced proliferation signature in this patient cohort. Subsequently the multi-gene signature score (Equation 2) was calculated and patients were either assigned to the low or high expression group. Patients in the low expression group are predicted to have good prognosis whereas patients in the high expression group are predicted to have poor prognosis. Disease-free survival was used as follow-up end point. Univariate and multivariate Cox proportional hazard ratio modeling followed by the Wald test was used to evaluate the reduced proliferation signature (R survival package version 2.36-2). For a subgroup of the cohort histological grade was unknown, median imputation was applied for those patients (R e1071 package v1.5-24). Moreover multivariate models with and without the signature were evaluated with the C-index (R survival package v2.36-2).

Table 4.1 The reduced proliferation signature

| Gene Symbol | Entrez Gene ID | Name |
|--------------------|-----------------------|--|
| BUB1B | 701 | Budding uninhibited by benzimidazoles 1 homolog beta (yeast) |
| CCNA2 | 890 | Cyclin A2 |
| CCNB2 | 9133 | Cyclin B2 |
| FANCI | 55215 | Fanconi anemia, complementation group I |
| MELK | 9833 | Maternal embryonic leucine zipper kinase |
| NCAPH | 23397 | Non-SMC condensing I complex, subunit H |
| RRM2 | 6241 | Ribonucleotide reductase M2 |
| SKA3 | 221150 | Spindle and kinetochore associated complex 3 |
| UBE2C | 11065 | Ubiquitin-conjugating enzyme E2C |
| UBE2T | 29089 | Ubiquitin-conjugating enzyme E2T (putative) |

Results

To reduce the number of genes in the proliferation signature, genes were ranked according to their individual prognostic power in each of the 5 datasets used in the original study (Supplementary Data File S1)¹⁶. After filtering and gene ranking, the final reduced proliferation signature consisted of ten genes, which is a reduction of 90% (Table 4.1).

The original basis of the proliferation signature was in gene expression studies carried out *in vitro*. To ensure that the remaining genes accurately reflect proliferation status *per se*, especially when assessed by qPCR, we evaluated the reduced signature both *in vitro* and *ex vivo*. First, five different cancer cell lines (MCF7, HeLa, HT-29, U-2 OS and DU145) were cultured in either normal or serum-starved conditions. Figure 4.1A shows that expression of the reduced proliferation signature was significantly lower upon serum starvation compared to control growing conditions ($P = 1.52 \times 10^{-11}$, t-test). Individual genes showed a similar pattern (Supplementary Figure S4.1).

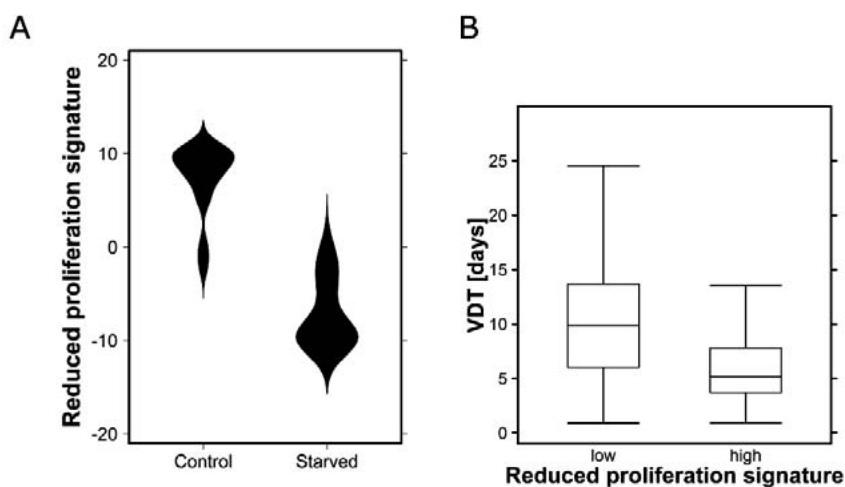


Figure 4.1

In vitro and *ex vivo* qPCR validation of the reduced proliferation signature. Difference in reduced proliferation signature score in normal versus starvation conditions were compared in a series of cell lines (A). A set of xenografts ($n = 180$) was dichotomized in a low and a high proliferation group based on the reduced proliferation signature and corresponding volume doubling times (VDTs) were compared (B).

We then assessed expression of the 10 genes in a panel of tumor xenografts with known volume doubling times (VDTs)³³⁻³⁶ originating from different cancer cell lines. Xenografts were assigned to either the low or high proliferation group based on expression of the reduced proliferation signature. Although proliferation rate is not expected to be the only parameter which influences gross tumor growth (e.g. rates of cell turnover are also important) we hypothesized that VDTs in the group with high proliferation should be reduced compared with xenografts with low proliferation. Figure 4.1B confirms this hypothesis: a significant difference in VDTs between high- and low-proliferation signature xenografts was observed ($P = 5.32 \times 10^{-6}$, fold-change $\left[\frac{\overline{\text{VDT}}^{\text{low proliferation}}}{\overline{\text{VDT}}^{\text{high proliferation}}} \right] = 1.60$; t-test).

To demonstrate its prognostic power, the reduced signature was evaluated in two large gene expression-based meta-datasets of 1820 breast and 862 NSCLC patients. None of these datasets were included in the original study; all were fully independent.

Patients were stratified based on the reduced proliferation signature and Cox proportional hazards modeling was used to assess performance. Patient classification with the reduced proliferation signature could stratify breast (Figure 4.2A: HR = 1.63; 95% CI 1.39 - 1.92; $P = 1.42 \times 10^{-9}$ Wald test) and NSCLC patients (Figure 4.2B: HR = 1.35; 95% CI 1.10 - 1.66; $P = 34.47 \times 10^{-3}$ Wald test) into groups with distinct prognostic profiles. High expression of the reduced proliferation signature correlated with poor survival in all patient groups. For NSCLC subgroup analyses were performed for the adenocarcinoma and squamous cell carcinoma patient groups, for these are significantly distinct disease states. NSCLC adenocarcinoma patients could be grouped into cohorts with significantly different survival properties (Figure 4.2C: HR = 1.64; 95% CI 1.30 – 2.06; $P = 3.01 \times 10^{-5}$ Wald test). However in the squamous cell carcinoma cohort the reduced proliferation signature had no prognostic power (Figure 4.2D: HR = 0.66; 95% CI 0.41 - 1.04; $P = 7.14 \times 10^{-2}$ Wald test). In Supplementary Figures S4.2 and S4.3 Kaplan-Meier survival curves for the individual datasets are provided. These data indicate that reduction of the proliferation signature was successful; the reduced signature could stratify patients into groups with significant differences in survival.

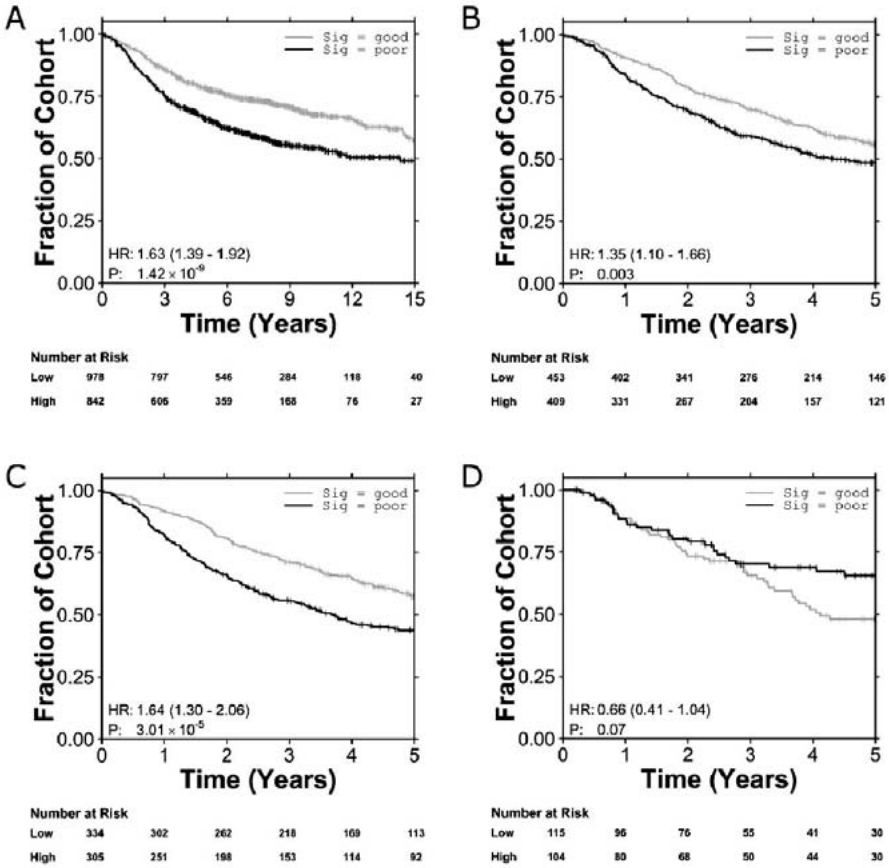


Figure 4.2

Validation of the reduced proliferation signature in a breast (A) and non-small cell lung (B) cancer meta-dataset. For NSCLC a subgroup analysis was performed (adenocarcinoma (C) and squamous cell carcinoma (D)). [HR: hazard ratio, P: p-value Wald test].

To confirm the prognostic performance of the reduced proliferation signature when evaluated by qPCR, we tested its performance in a further independent cohort of 129 lymph-node negative breast cancer patients. This patient group is distinct from those used for model development and from those in the meta-dataset analysis. Table 4.2 displays patient and treatment characteristics. The reduced proliferation signature stratified the cohort into groups predicted to have either good (low proliferation) or poor prognosis (high proliferation). Figure 4.3A shows that the patient group predicted to have poor prognosis had significantly worse disease-specific survival than the good prognosis group (HR = 2.25; 95% CI 1.01 - 4.99; P =

4.60×10^{-2} Wald test). The majority of this cohort were stage I patients, therefore a subgroup analysis was performed. The reduced proliferation signature could stratify stage I patients in two groups with highly significant differences in prognosis (Figure 4.3B: HR = 5.92; 95% CI 1.62 - 21.59; $P = 7.03 \times 10^{-3}$ Wald test).

To investigate whether the signature's prognostic power was independent of other clinical factors multivariate Cox proportional hazard ratio modeling was used. In the whole patient cohort the reduced signature performed comparable to stage (Supplementary Table S4.1), however it did, like the other factors included, not reach statistical significance (HR = 1.73; 95% CI 0.73 - 4.12; $P = 0.215$ Wald test). In stage I patients the reduced proliferation signature was the top prognostic factor (Table 4.3, HR = 7.23; 95% CI 1.65 - 31.95; $P = 8.57 \times 10^{-3}$ Wald test). C-indexes for multivariate Cox proportional hazard ratio models of clinical parameters with and without the reduced proliferation signature were calculated. Both in the whole cohort and stage I patient group adding the signature increased prognostic power, in the stage I patient group the signature alone outperformed the model comprising clinical parameters (Supplementary Table S4.2).

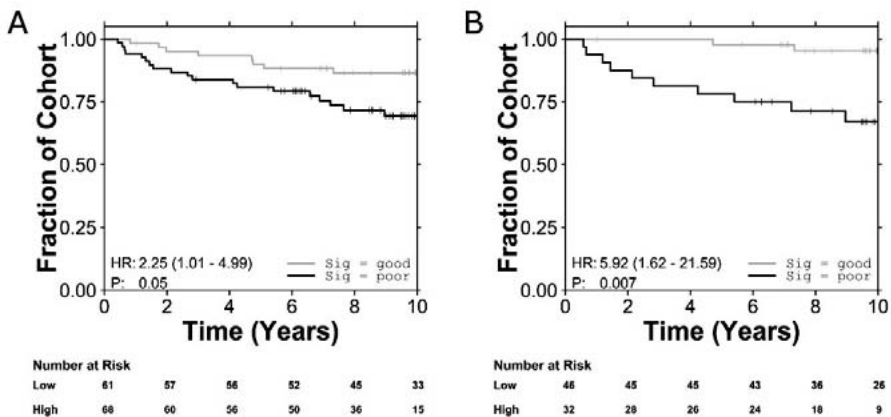


Figure 4.3

Validation of the reduced proliferation signature with qPCR in a breast cancer patient cohort; for the whole patient group (A) and only stage I patients (B), respectively. [HR: hazard ratio, P: p-value Wald test].

Table 4.2 Baseline demographics of breast cancer patients in low and high risk groups assessed with the reduced proliferation signature (full characteristics were presented previously⁵³)

| Factor | All patients (N = 129) | Low risk (N = 61) | High risk (N = 68) | P* |
|---------------------------|---------------------------|----------------------|-----------------------|---------|
| Surgery | | | | 0.371 |
| Mastectomy | 55 | 23 | 32 | |
| Lumpectomy | 74 | 38 | 36 | |
| Radiotherapy | | | | 0.935 |
| Yes | 84 | 39 | 45 | |
| No | 45 | 22 | 23 | |
| Stage of disease | | | | 0.004 |
| I | 78 | 46 | 32 | |
| II | 48 | 14 | 34 | |
| III | 3 | 1 | 2 | |
| Grade⁺ | | | | 0.142** |
| G1 | 9 | 5 | 4 | |
| G2 | 48 | 25 | 23 | |
| G3 | 40 | 13 | 27 | |
| Not recorded | 32 | 18 | 14 | |
| Histological type | | | | 0.017 |
| Ductal | 90 | 45 | 45 | |
| Lobular | 19 | 12 | 7 | |
| Other | 20 | 4 | 16 | |
| Menopausal | | | | 0.928 |
| Pre | 26 | 13 | 13 | |
| Post | 103 | 48 | 55 | |
| Treatment | | | | 0.121 |
| Lumpectomy + radiotherapy | 74 | 38 | 36 | |
| Mastectomy | 44 | 21 | 23 | |
| Mastectomy + radiotherapy | 11 | 2 | 9 | |

* p value χ^2 -test

** without including those with missing or unknown values

+ grading was performed according to Bloom and Richardson, by the method modified by Elston and Ellis⁵⁴

Table 4.3 Results multi-variate Cox regression model stage I patient group (78 patients)

| Factor | HR | CI | p |
|---|------|-------------|-------|
| Reduced proliferation signature | 7.23 | 1.65 - 31.6 | 0.009 |
| Grade (vs. grade 1) | | | |
| Grade 2 | 0.79 | 0.14 - 4.36 | 0.783 |
| Grade 3 | 0.17 | 0.02 - 1.36 | 0.095 |
| Histological type (vs. ductal) | | | |
| Histological type lobular | 0 | 0.00 - inf | 0.998 |
| Histological type other | 2.02 | 0.54 - 7.64 | 0.298 |
| Age | 0.99 | 0.90 - 1.11 | 0.992 |
| Menopausal (pre vs. post) | 0.95 | 0.06 - 14.7 | 0.970 |
| Treatment (vs. lumpectomy + radiotherapy) | | | |
| Mastectomy | 2.99 | 0.79 - 11.3 | 0.106 |

Discussion

We previously reported a microarray-based proliferation signature with high prognostic power in several large microarray datasets encompassing different cancer types. Here, we successfully reduced the number of genes in the proliferation signature to a more appropriate scale for low throughput technologies. This could greatly facilitate the translation into a clinically applicable test²⁵. In two large independent gene expression meta-datasets for breast and NSCLC the reduced signature separated the patients into groups with significant distinct survival properties.

A subgroup analysis for the NSCLC cohort showed high prognostic power in adenocarcinoma patients, whereas in squamous cell carcinoma patients no prognostic power was observed. Earlier studies have shown similar data for other measures of proliferation; high proliferation was significantly associated with incidence of metastasis and worse survival in adenocarcinomas, but not in squamous cell carcinomas^{55,56}. Recapitulating decreasing the number of signature genes resulted in a new marker with high performance across different cancer types.

Several genes in the signature have previously been implicated in cancer outcome^{14,57-60}. UBE2C (ubiquitin-conjugating enzyme E2C) expression was correlated with malignant progression in thyroid carcinomas and demonstrated prognostic power in ovarian cancer⁶¹⁻⁶³, in which high expression was associated with worse survival. Over-expression of RRM2 (ribonucleotide reductase M2) showed association with chemotherapy resistance⁶⁴. Furthermore, a large fraction

of the published gene expression signatures include clusters of proliferation-associated genes and several of the reduced proliferation genes are represented in these clusters^{1,12,65,66}.

As a last step the reduced proliferation signature was evaluated with qPCR in an independent breast cancer patient cohort. This patient group consisted entirely of patients without axillary lymph node metastases, and who did not receive systemic adjuvant therapy, making it possible to distinguish a pure prognostic value of the proliferation signature. The reduced proliferation signature stratified patients into groups with different survival properties and showed high prognostic power especially in stage I patients. A high disease-specific survival was observed in the stage I patients identified as having low risk. This suggests the reduced proliferation signature might be useful in identifying high risk stage I breast cancer patients that could benefit from additional therapy like chemo-radiation or accelerated radiotherapy, whereas the low risk group would not.

Currently two large prospective trials have been started to address the predictive performance of two gene expression signatures in early breast cancer^{22,23}. Both these signatures include a subset of proliferation genes and several meta-analyses show evidence that the prognostic value of these signatures is mostly attributed to this process^{12,15}. Therefore a signature reflecting merely proliferation could make its interpretation easier. Furthermore the prognostic power of the reduced proliferation signature was not limited to breast cancer; it also had a high performance in a NSCLC adenocarcinoma meta-dataset.

Thus, we here show that the array based proliferation signature could be reduced to 10 genes. This reduced proliferation signature can be applied in small tissue samples, including possibly FFPE material, adding to its clinical applicability. The pure prognostic power of the signature was validated in an independent breast cancer patient cohort, where it was shown to be particularly useful to select patients that would benefit from more aggressive therapy. To fully grasp the potential prognostic or predictive role of the signature it further should be tested in prospective trials and translated from a relative to an absolute measure.

Acknowledgements

This work was supported by the Center for Translational Molecular Medicine (www.ctmm.nl) (AIRFORCE Project Ref. 030-103) and the Ontario Institute for Cancer Research through funding provided by the Government of Ontario to PCB.

Supplementary figures and tables

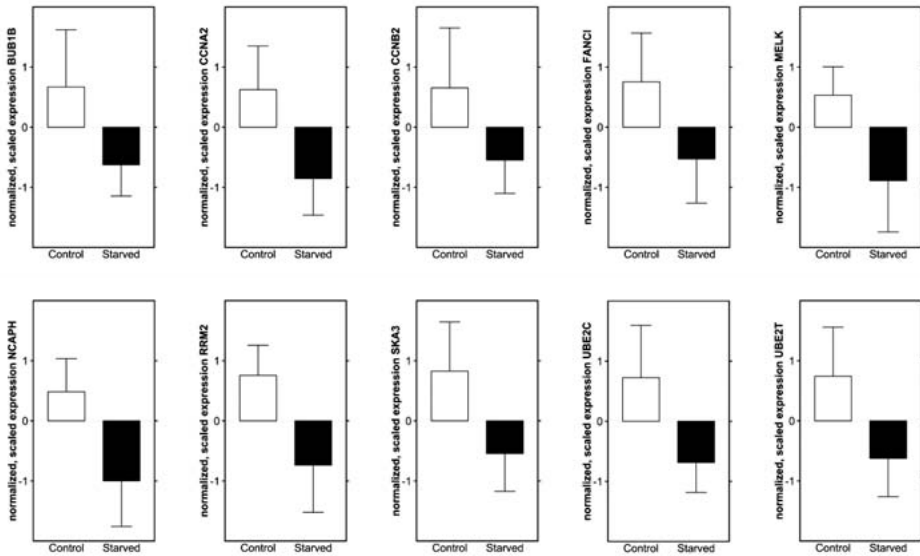


Figure S4.1

In vitro validation of the individual genes in the reduced proliferation signature. Difference in normalized, scaled expression between normal and starvation conditions in a panel of cell lines are shown.

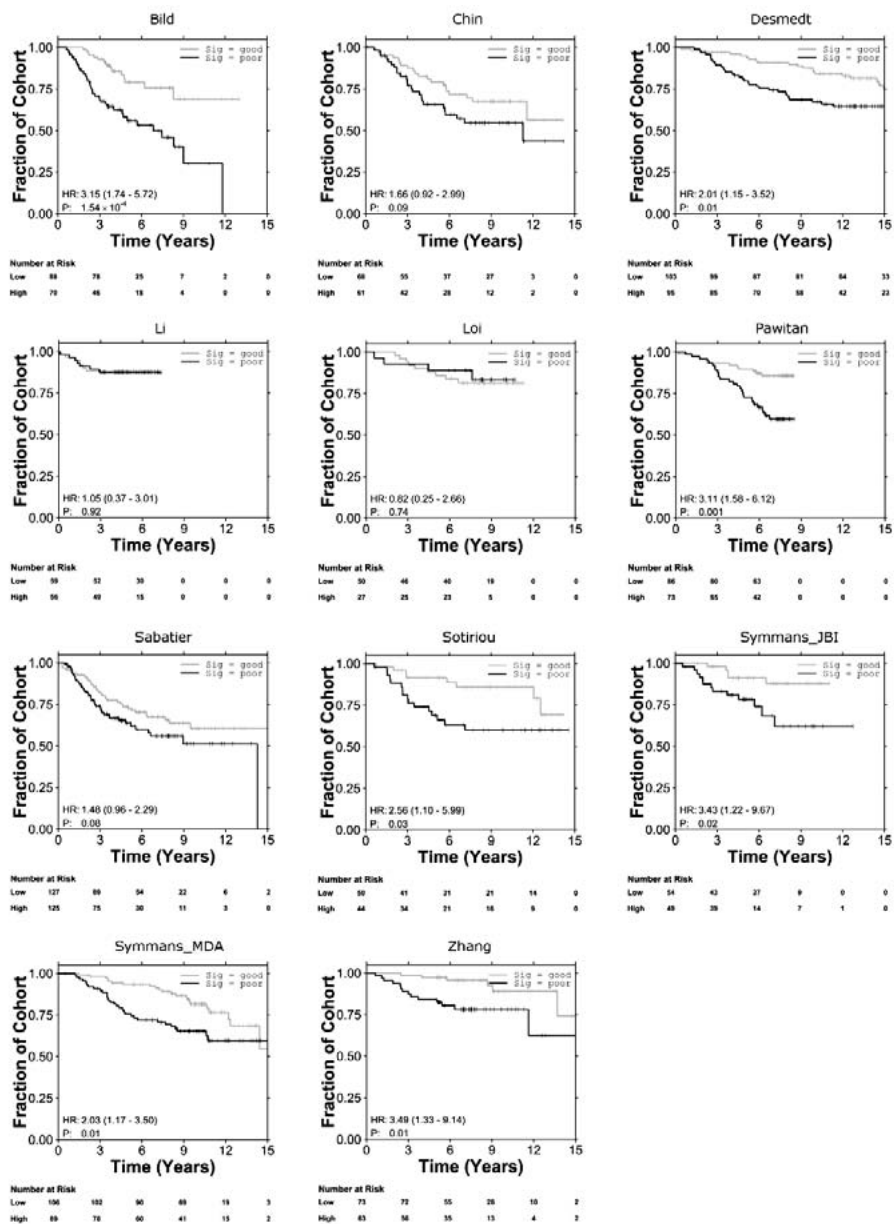


Figure S4.2

Validation of the reduced proliferation signature in the individual breast cancer datasets used for meta-dataset analysis (HR: hazard ratio, p: p-value Wald test).

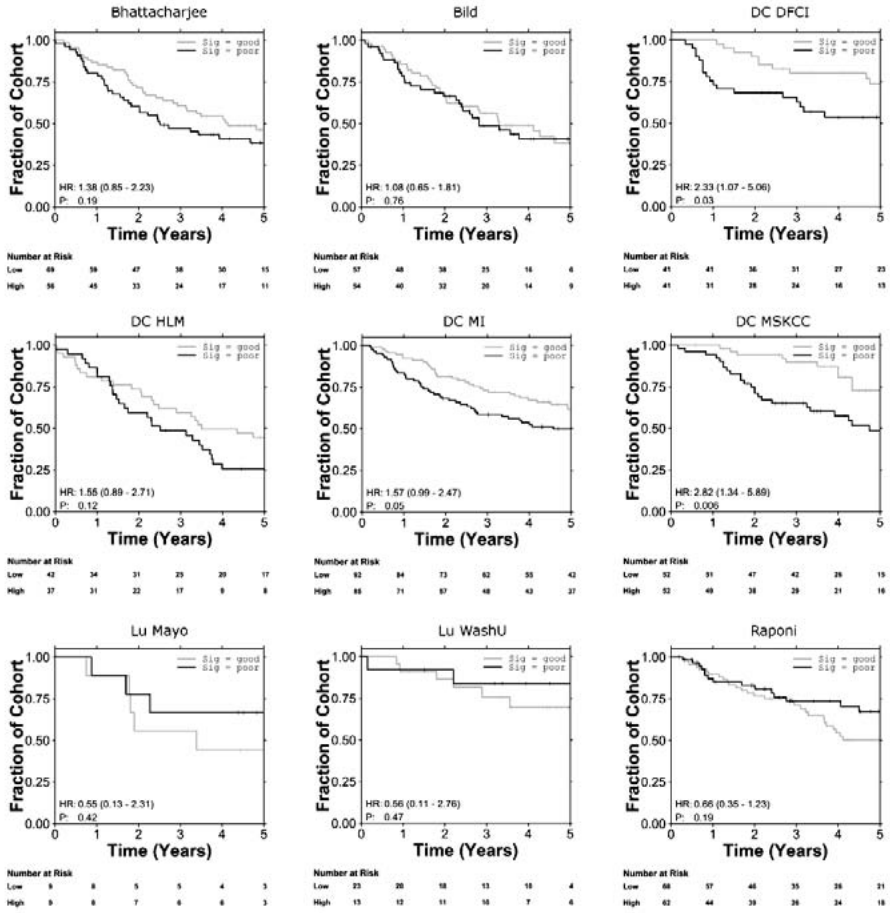


Figure S4.3

Validation of the reduced proliferation signature in the individual non-small cell lung cancer datasets used for meta-dataset analysis (HR: hazard ratio, p: p-value Wald test).

Chapter 4

Table S4.1 Results multi-variate Cox regression model (129 patients)

| Factor | HR | CI | p |
|---|------|-------------|-------|
| Reduced proliferation signature | 1.73 | 0.73 - 4.12 | 0.215 |
| Stage (vs. stage I) | | | |
| Stage II/III | 1.75 | 0.76 - 4.00 | 0.185 |
| Grade (vs. grade 1) | | | |
| Grade 2 | 0.71 | 0.15 - 3.34 | 0.667 |
| Grade 3 | 0.70 | 0.14 - 3.48 | 0.663 |
| Histological type (vs. ductal) | | | |
| Histological type lobular | 0.94 | 0.26 - 3.37 | 0.930 |
| Histological type other | 1.90 | 0.79 - 4.58 | 0.154 |
| Age | 0.99 | 0.95 - 1.04 | 0.663 |
| Menopausal (pre vs. post) | 0.52 | 0.13 - 2.12 | 0.359 |
| Treatment (vs. lumpectomy + radiotherapy) | | | |
| Mastectomy | 1.49 | 0.63 - 3.55 | 0.367 |
| Mastectomy + radiotherapy | 0.66 | 0.13 - 3.29 | 0.616 |

Table S4.2 Results C-index calculation

| Model | All patients | Stage I patients |
|--|--------------|------------------|
| Signature | 0.603 | 0.735 |
| Clinical parameters | 0.674 | 0.692 |
| Clinical parameters + signature | 0.686 | 0.826 |
| Clinical parameters (without histological grade) | 0.668 | 0.656 |
| Clinical parameters (without histological grade) + signature | 0.682 | 0.787 |

References

1. Shedden, K., *et al.* Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* **14**, 822-827 (2008).
2. Gomez-Raposo, C., Mendiola, M., Barriuso, J., Hardisson, D. & Redondo, A. Molecular characterization of ovarian cancer by gene-expression profiling. *Gynecol Oncol* **118**, 88-92 (2010).
3. Sotiriou, C. & Pusztai, L. Gene-expression signatures in breast cancer. *N Engl J Med* **360**, 790-800 (2009).
4. Oberthuer, A., *et al.* Prognostic impact of gene expression-based classification for neuroblastoma. *J Clin Oncol* **28**, 3506-3515 (2010).
5. Lau, S.K., *et al.* Three-gene prognostic classifier for early-stage non small-cell lung cancer. *J Clin Oncol* **25**, 5562-5569 (2007).
6. Chen, H.Y., *et al.* A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* **356**, 11-20 (2007).
7. Fan, C., *et al.* Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* **355**, 560-569 (2006).
8. Ein-Dor, L., Zuk, O. & Domany, E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci U S A* **103**, 5923-5928 (2006).
9. Kuo, W.P., Jenssen, T.K., Butte, A.J., Ohno-Machado, L. & Kohane, I.S. Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* **18**, 405-412 (2002).
10. Tan, P.K., *et al.* Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* **31**, 5676-5684 (2003).
11. Haibe-Kains, B., Desmedt, C., Sotiriou, C. & Bontempi, G. A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? *Bioinformatics* **24**, 2200-2208 (2008).
12. Weigelt, B., Baehner, F.L. & Reis-Filho, J.S. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. *J Pathol* **220**, 263-280 (2010).
13. Desmedt, C., *et al.* Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res* **14**, 5158-5165 (2008).
14. Whitfield, M.L., George, L.K., Grant, G.D. & Perou, C.M. Common markers of proliferation. *Nat Rev Cancer* **6**, 99-106 (2006).
15. Wirapati, P., *et al.* Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* **10**, R65 (2008).
16. Starmans, M.H., *et al.* Robust prognostic value of a knowledge-based proliferation signature across large patient microarray studies spanning different cancer types. *Br J Cancer* **99**, 1884-1890 (2008).
17. Michiels, S., Koscielny, S. & Hill, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* **365**, 488-492 (2005).
18. Dupuy, A. & Simon, R.M. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* **99**, 147-157 (2007).
19. Boulesteix, A.L. & Slawski, M. Stability and aggregation of ranked gene lists. *Brief Bioinform* **10**, 556-568 (2009).

20. Subramanian, J. & Simon, R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Natl Cancer Inst* **102**, 464-474 (2010).
21. Boutros, P.C., *et al.* Re: Gene Expression-Based Prognostic Signatures in Lung Cancer: Ready for Clinical Use? *J Natl Cancer Inst* (2010).
22. Sparano, J.A. TAILORx: trial assigning individualized options for treatment (Rx). *Clin Breast Cancer* **7**, 347-350 (2006).
23. Bogaerts, J., *et al.* Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. *Nat Clin Pract Oncol* **3**, 540-551 (2006).
24. Zien, A., Fluck, J., Zimmer, R. & Lengauer, T. Microarrays: how many do you need? *J Comput Biol* **10**, 653-667 (2003).
25. Zhou, Y.H., Raj, V.R., Siegel, E. & Yu, L. Standardization of Gene Expression Quantification by Absolute Real-Time qRT-PCR System Using a Single Standard for Marker and Reference Genes. *Biomark Insights* **5**, 79-85 (2010).
26. Expression profiling—best practices for data generation and interpretation in clinical trials. *Nat Rev Genet* **5**, 229-237 (2004).
27. Miller, L.D., *et al.* An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A* **102**, 13550-13555 (2005).
28. Wang, Y., *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365**, 671-679 (2005).
29. van de Vijver, M.J., *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347**, 1999-2009 (2002).
30. Zhao, H., *et al.* Gene expression profiling predicts survival in conventional renal cell carcinoma. *PLoS Med* **3**, e13 (2006).
31. Beer, D.G., *et al.* Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* **8**, 816-824 (2002).
32. Barsyte-Lovejoy, D., *et al.* The c-Myc oncogene directly induces the H19 noncoding RNA by allele-specific binding to potentiate tumorigenesis. *Cancer Res* **66**, 5330-5337 (2006).
33. Oostendorp, M., *et al.* Quantitative molecular magnetic resonance imaging of tumor angiogenesis using cNGR-labeled paramagnetic quantum dots. *Cancer Res* **68**, 7676-7683 (2008).
34. Dubois, L., *et al.* Imaging of CA IX with fluorescent labelled sulfonamides distinguishes hypoxic and (re)-oxygenated cells in a xenograft tumour model. *Radiother Oncol* **92**, 423-428 (2009).
35. Dubois, L., *et al.* Inhibition of 4E-BP1 sensitizes U87 glioblastoma xenograft tumors to irradiation by decreasing hypoxia tolerance. *Int J Radiat Oncol Biol Phys* **73**, 1219-1227 (2009).
36. Rouschop, K.M., *et al.* The unfolded protein response protects human tumor cells during hypoxia through regulation of the autophagy genes MAP1LC3B and ATG5. *J Clin Invest* **120**, 127-141 (2010).
37. Bild, A.H., *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353-357 (2006).
38. Chin, K., *et al.* Genomic and transcriptional aberrations linked to breast cancer pathophysiology. *Cancer Cell* **10**, 529-541 (2006).
39. Desmedt, C., *et al.* Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res* **13**, 3207-3214 (2007).

40. Loi, S., *et al.* Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. *BMC Genomics* **9**, 239 (2008).
41. Pawitan, Y., *et al.* Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* **7**, R953-964 (2005).
42. Sotiriou, C., *et al.* Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* **98**, 262-272 (2006).
43. Bos, P.D., *et al.* Genes that mediate breast cancer metastasis to the brain. *Nature* **459**, 1005-1009 (2009).
44. Li, Y., *et al.* Amplification of LAPT4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. *Nat Med* **16**, 214-218 (2010).
45. Sabatier, R., *et al.* A gene expression signature identifies two prognostic subgroups of basal breast cancer. *Breast Cancer Res Treat* (2010).
46. Symmans, W.F., *et al.* Genomic index of sensitivity to endocrine therapy for breast cancer. *J Clin Oncol* **28**, 4111-4119 (2010).
47. Zhang, Y., *et al.* The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. *Breast Cancer Res Treat* **116**, 303-309 (2009).
48. Raponi, M., *et al.* Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res* **66**, 7466-7472 (2006).
49. Bhattacharjee, A., *et al.* Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* **98**, 13790-13795 (2001).
50. Lu, T.P., *et al.* Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. *Cancer Epidemiol Biomarkers Prev* **19**, 2590-2597 (2010).
51. Irizarry, R.A., *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-264 (2003).
52. Dai, M., *et al.* Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res* **33**, e175 (2005).
53. Span, P.N., *et al.* Mammaglobin is associated with low-grade, steroid receptor-positive breast tumors from postmenopausal patients, and has independent prognostic value for relapse-free survival time. *J Clin Oncol* **22**, 691-698 (2004).
54. Elston, C.W. & Ellis, I.O. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* **19**, 403-410 (1991).
55. Komaki, R., *et al.* Apoptosis and mitosis as prognostic factors in pathologically staged N1 nonsmall cell lung cancer. *Int J Radiat Oncol Biol Phys* **36**, 601-605 (1996).
56. Hommura, F., *et al.* Prognostic significance of p27KIP1 protein and ki-67 growth fraction in non-small cell lung cancers. *Clin Cancer Res* **6**, 4073-4081 (2000).
57. Hao, J., *et al.* Elevated expression of UBE2T in lung cancer tumors and cell lines. *Tumour Biol* **29**, 195-203 (2008).
58. Glinsky, G.V. Genomic models of metastatic cancer: functional analysis of death-from-cancer signature genes reveals aneuploid, anoikis-resistant, metastasis-enabling phenotype with altered cell cycle control and activated Polycomb Group (PcG) protein chromatin silencing pathway. *Cell Cycle* **5**, 1208-1216 (2006).

59. Marie, S.K., *et al.* Maternal embryonic leucine zipper kinase transcript abundance correlates with malignancy grade in human astrocytomas. *Int J Cancer* **122**, 807-815 (2008).
60. Ryu, B., Kim, D.S., Deluca, A.M. & Alani, R.M. Comprehensive expression profiling of tumor cell lines identifies molecular signatures of melanoma progression. *PLoS One* **2**, e594 (2007).
61. Berlingieri, M.T., *et al.* Ubch10 expression may be a useful tool in the prognosis of ovarian carcinomas. *Oncogene* **26**, 2136-2140 (2007).
62. Pallante, P., *et al.* Ubch10 overexpression may represent a marker of anaplastic thyroid carcinomas. *Br J Cancer* **93**, 464-471 (2005).
63. van Ree, J.H., Jeganathan, K.B., Malureanu, L. & van Deursen, J.M. Overexpression of the E2 ubiquitin-conjugating enzyme Ubch10 causes chromosome missegregation and tumor formation. *J Cell Biol* **188**, 83-100 (2010).
64. Boukovinas, I., *et al.* Tumor BRCA1, RRM1 and RRM2 mRNA expression levels and clinical response to first-line gemcitabine plus docetaxel in non-small-cell lung cancer patients. *PLoS One* **3**, e3695 (2008).
65. Whitfield, M.L., *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* **13**, 1977-2000 (2002).
66. Dai, H., *et al.* A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. *Cancer Res* **65**, 4059-4066 (2005).

Chapter 5

A simple but highly effective approach to evaluate the prognostic performance of gene expression signatures

Under consideration at PLoS one

Maud H.W. Starmans, Glenn Fung, Harald Steck, Bradly G. Wouters, Philippe Lambin

Abstract

Background: Highly parallel analysis of gene expression has recently been used to identify gene sets or 'signatures' to improve patient diagnosis and risk stratification. Once a signature is generated, traditional statistical testing is used to evaluate its prognostic performance. However, due to the dimensionality of microarrays, this can lead to false interpretation of these signatures.

Principle findings: A method was developed to test batches of a user-specified number of randomly chosen signatures in patient microarray datasets. The percentage of random generated signatures yielding prognostic value was assessed using ROC analysis by calculating the area under the curve (AUC) in six public available cancer patient microarray datasets.

We found that a signature consisting of randomly selected genes has an average 10% chance of reaching significance when assessed in a single dataset, but can range from 1% to ~40% depending on the dataset in question. Increasing the number of validation datasets markedly reduces this number.

Conclusions: We have shown that the use of an arbitrary cut-off value for evaluation of signature significance is not suitable for this type of research, but should be defined for each dataset separately. Our method can be used to establish and evaluate signature performance of any derived gene signature in a dataset by comparing its performance to thousands of randomly generated signatures. It will be of most interest for cases where few data are available and testing in multiple datasets is limited.

Introduction

In recent years, DNA microarray technology has been increasingly used in oncology. It has provided insight into the biological mechanisms underlying tumour formation and identified new therapy targets^{1,2}. However, most studies performed in this field identify gene sets, or so-called signatures, which can be used to improve diagnosis and risk stratification³⁻⁶. These signatures can be acquired through supervised analysis methods⁷. Both patient microarray and clinical data are directly used to find the genes that correlate with tumour type or patient outcome⁸⁻¹². Also biology-based signatures can be used for patient prognosis, which are usually derived from *in vitro* microarray data^{2,13-15}. Though the performance of these classifiers can be very high in the dataset studied, application of these signatures in other datasets is often limited and data reproduction is not straightforward¹⁶. Furthermore, signatures identified in comparable studies show little overlap in gene content^{1,10,17-20}. Michiels *et al.*⁴ showed that identified gene lists were highly variable within one dataset and depended on the patients included in the training set. Further, they demonstrated that several published gene classifiers did not classify patients better than by chance. They stress that validation is an important issue in microarray research. Fan *et al.*²¹ repeated and extended these analyses 5 years later and made similar conclusions. Moreover Boutros *et al.*¹⁹ amongst other showed that the use of different statistical procedures could identify multiple highly prognostic signatures from one dataset^{22,23}. An extensive analysis of the effect of different statistics on ranked gene lists showed large variability²⁴.

A major challenge with DNA microarray technology is to take account of variability across a very large number of parameters¹. This variability arises from several sources: the biological samples, hybridisation protocols, scanning, and image and statistical analysis⁷. In a recent review, Dupuy *et al.*¹ demonstrated that proper methodology in pre-processing and statistical analysis is essential in these sorts of studies. They found that a large subset of published microarray studies show flaws in the applied analysis; serious mistakes are made in the selection of genes and inadequate control of multiple testing is performed. The issue of multiple testing is crucial, as microarrays monitor the expression of thousands of genes, while the number of samples is relatively small.

Statistical significance of the differences in gene expression patterns for different patient groups or tumour types is often determined with traditional statistical

testing procedures, such as the two-sample t-tests or Wilcoxon rank sum tests^{1,7,20}. These procedures are challenged with serious multiplicity and without employment of a correction for multiple testing, the number of false positives will be extremely high. Various methods have been developed to overcome this problem of identifying differentially expressed genes and are used to create gene signatures^{11,19,20,25}.

More importantly multiple testing is often not considered in evaluating the prognostic power of signatures. Once a signature is created, its prognostic power is determined with traditional survival statistics and standard cut-off values for significance. We hypothesise that this can lead to high numbers of false prognostic signatures when the number of evaluated datasets is limited. Therefore we sought to develop a simple method to take into account the high-dimensionality of microarrays in the phase of evaluating signature prognosticity.

To quantify the problem of multiple testing we have developed a method to test batches of random signatures in microarray datasets. We show that the average chance that a random signature produces a prognostic result in one dataset is approximately 10% but can range from 1% to ~40%. Increasing the number of datasets reduces this false positive rate significantly. As a result of this high degree of variability amongst datasets, we developed a method that can be used to determine an appropriate threshold level of significance that must be reached for a given signature. This is done by testing a set of randomly chosen signatures along with the signature of interest within the dataset under investigation.

Materials and methods

Random signature testing

A method to test the prognostic value of random gene signatures of a predefined size on a microarray dataset was developed in Matlab (Matlab 7.1, The Mathworks, Natick, MA, USA). Unless indicated otherwise, analyses were performed using this program. The program creates a user-specified number of random gene sets, consisting of a user-specified number of genes. For a given dataset all genes on the respective microarray were used to create the random signatures. This batch of random signatures was then tested on a dataset by means of a signature score calculation.

Datasets

Patient microarray and clinical follow-up data were collated to test the random gene sets. Datasets are publicly available in the microarray databases Gene Expression Omnibus (GEO: <http://www.ncbi.nlm.nih.gov/projects/geo/>) and Stanford Microarray Database (SMD: <http://genome-www.stanford.edu/microarray>) and elsewhere. Accessory clinical and follow-up data were also given or provided by the authors on request. Table 5.1 provides an overview of the datasets and the databases, where these are accessible. Data filtering and pre-processing is explained in the supplementary data.

Signature score calculation

Expression data of the genes in a signature was extracted from the dataset. The following step was used to calculate a signature score for each patient included in the dataset. This score was defined as the average expression value of the genes in the signature (Equation 1). When a gene was represented by more than one reporter on an array, the expression of the reporters was averaged before signature calculation. The signature scores for each patient were then coupled to the survival data of the patients. The signature score was used to median dichotomize the patient cohorts.

$$\text{Score} = \frac{\sum_{i=1}^N \text{exp}_{i,m}}{N} \quad \text{Equation (1)}$$

Where: Score, signature score; N, number of genes in the signature; $\text{exp}_{i,m}$, gene expression of gene i in sample m .

In a second setup expression of the genes in the signature were used for K-nearest neighbor classification (KNN) combined with leave-one-out-cross validation (LOOCV). With this method one patient is withheld and the class membership of this patient is predicted using the KNN model (`knnclassify` function in Matlab) built on the remaining patients. The event parameter of the survival data was used as training class. This procedure was repeated for each patient, resulting in a class prediction for the whole cohort.

Table 5.1 Overview of the analysed patient microarray datasets.

| Dataset | Cancer type | # patients with survival data | # Unigenes on array | Average # reporters measured per UnigeneID | Source |
|---------------|---------------|-------------------------------|---------------------|--|---|
| Miller | Breast cancer | 236 | 20,647 | 1.97 | GEO accession GSE3494: http://www.ncbi.nlm.nih.gov/projects/geo/ |
| Wang | Breast cancer | 286 | 12,867 | 1.61 | GEO accession GSE2034: http://www.ncbi.nlm.nih.gov/projects/geo/ |
| van de Vijver | Breast cancer | 295 | 18,781 | 1.21 | http://microarray-pubs.stanford.edu/wound_NKI/ |
| Zhao | Renal cancer | 177 | 5,640 | 1.40 | SMD: http://smd.stanford.edu/ |
| Beer | Lung cancer | 86 | 5,396 | 1.15 | http://dot.ped.med.umich.edu:2000/ourimage/pub/Lung/index.html |
| Garber | Lung cancer | 24 | 4,936 | 1.15 | SMD: http://smd.stanford.edu/ |

Analysis

The signature scores, median dichotomized groups or KNN classifications were evaluated with the area under the curve (AUC) of the receiver operator curve (ROC). Definitions for AUC calculations are as follows:

- True positive: patient in the high score group that died from disease
- False positive: patient in the high score group that is alive
- True negative: patient in the low score group that is alive
- False negative: patient in the low score group that died from disease

A signature score was considered prognostic when the AUC is ≤ 0.4 or ≥ 0.6 . This cut-off value was based on the AUCs of several published gene signatures evaluated in the study of Ntzani et al.²⁶ (further details are given in the results section and supplementary data).

Results

In order to assess the potential for identifying prognostic gene signatures by chance alone in microarray based datasets a method was developed to test the prognostic value of batches of randomly generated signatures. Six different publicly available microarray datasets with follow-up data were used (Table 5.1). These six datasets

differ in number of patients, number of measured genes, number of reporters measured per gene, as well as platform and type of cancer.

For each dataset separately 5 batches of 10,000 random signatures were generated and tested. In each batch the number of genes (UnigeneIDs) in a gene set was predefined. The number of genes (UnigeneIDs) in the five batches were 10, 25, 50, 100 and 200 respectively. For example, the first batch included 10,000 random signatures, each consisting of ten genes. For each signature a patient score was derived, defined as the average of the expression of the genes in a signature (Equation 1). Each signature score was then tested for prognostic value by ROC analysis and determination of the AUC. Figure 5.1A-F shows the distribution of the AUCs for the first batch of 10,000 random signatures for the different datasets.

To define a reasonable cut-off value for the AUC values, we first searched for AUCs used in published gene signatures. However, the majority of studies do not evaluate gene signatures using the AUC. Most gene signatures are evaluated with Kaplan-Meier survival curves and log-rank tests. Kaplan-Meier survival analyses and ROC analyses are linked; a high AUC corresponds to a good separation in distinct survival groups. To be able to define a cut-off, we calculated the AUCs for the different gene sets as evaluated in the review by Ntzani *et al.*²⁶. The calculated AUCs as well as additional information are provided in the supplementary data. Based on these calculations we chose the cut-off values $AUC \leq 0.4$ and ≥ 0.6 .

In Figure 5.1G the percentage of signatures that passed the criteria for the different batches of signatures are given. These percentages range from 1% to ~40%, dependent on the dataset and the number of genes in the signatures. Table 5.2 provides the average, standard deviation as well as maximum and minimum AUC for the analyses with the gene sets consisting of ten genes. These data show that the larger the standard deviation, the higher the chance that a randomly generated signature is considered prognostic. Further, the maximum and minimum AUC show that very high signature performances can be found at random.

Table 5.2 A batch of 10,000 random signatures of 10 genes was tested in the six datasets.

| Dataset | Average (\pm standard deviation) AUC | Maximum AUC | Minimum AUC |
|---------------|---|-------------|-------------|
| Miller | 0.505 \pm 0.054 | 0.692 | 0.312 |
| Wang | 0.475 \pm 0.043 | 0.644 | 0.329 |
| van de Vijver | 0.499 \pm 0.073 | 0.744 | 0.297 |
| Zhao | 0.472 \pm 0.048 | 0.666 | 0.333 |
| Beer | 0.502 \pm 0.073 | 0.753 | 0.249 |
| Garber | 0.536 \pm 0.118 | 0.938 | 0.031 |

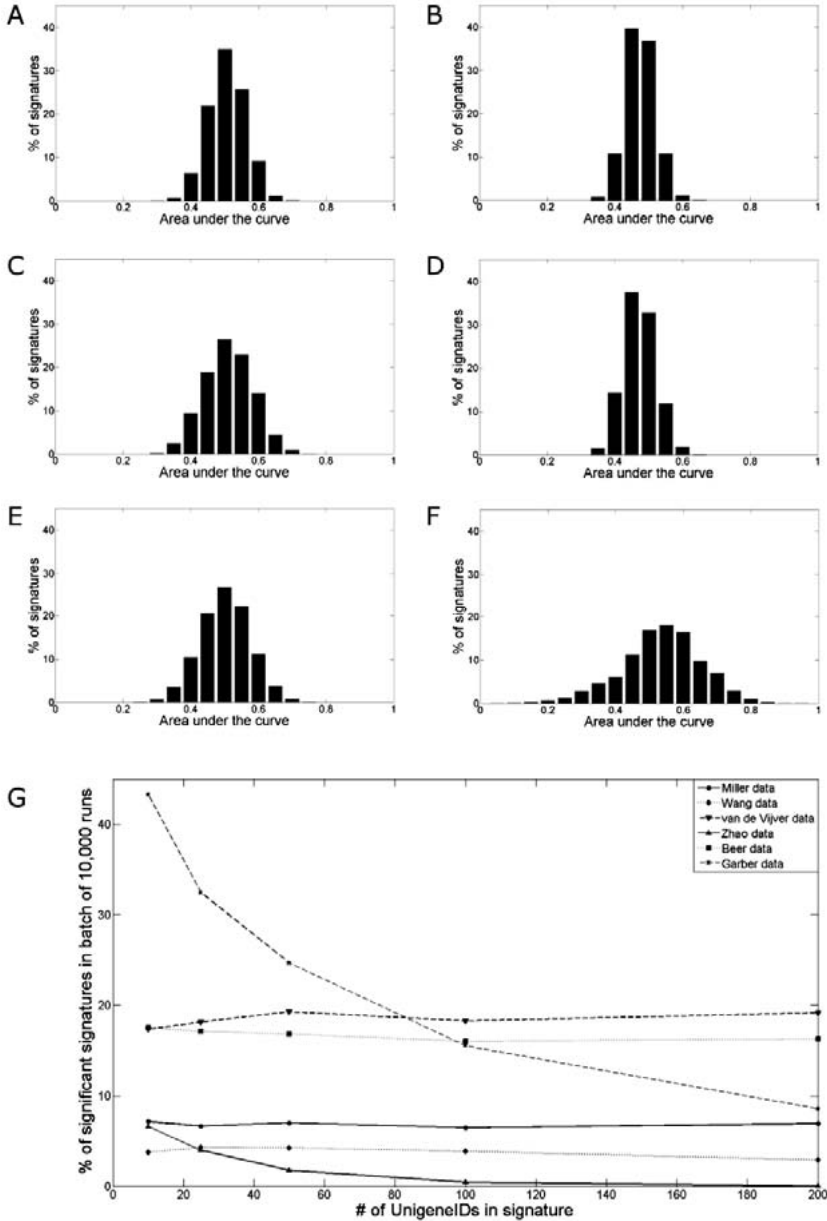


Figure 5.1

Random signature AUC distribution in different published microarray datasets (A: Miller, B: Wang, C: van de Vijver, D: Zhao, E: Beer and F: Garber dataset) for batches of 10,000 random gene sets with 10 genes evaluated with receiver-operator curves. The percentages of random gene sets passing the AUC criteria $AUC \leq 0.4$ or $AUC \geq 0.6$ for different signature sizes are displayed in G.

Table 5.3 A batch of 1,000,000 random signatures of 10 genes was tested in the six datasets.

| Dataset | Average (\pm standard deviation) AUC | Maximum AUC | Minimum AUC |
|---------------|---|-------------|-------------|
| Miller | 0.503 \pm 0.044 | 0.688 | 0.290 |
| Wang | 0.475 \pm 0.042 | 0.691 | 0.303 |
| van de Vijver | 0.511 \pm 0.072 | 0.778 | 0.209 |
| Zhao | 0.472 \pm 0.048 | 0.699 | 0.283 |
| Beer | 0.503 \pm 0.073 | 0.831 | 0.196 |
| Garber | 0.539 \pm 0.117 | 1.000 | 0.000 |

Sampling 10,000 gene sets is a small number compared to the total number of possible gene sets. In order to show that the 10,000 random gene sets are sufficient to estimate the AUC distribution, we tested batches of 1,000,000 signatures consisting of 10 genes in the six datasets. The AUC distributions for this permutation study were similar to the distributions for the batches of 10,000 gene sets (Table 5.3).

From the differences between the six datasets (Table 5.1), it could be that the number of patients, the number of genes (UnigeneIDs) and the number of reporters measured per gene influence the probability that a randomly chosen signature is considered prognostic. To further investigate the impact of these parameters, the Miller dataset was used. To determine the influence of patient number, the dataset was split in halves and in quarters. For these partial datasets the same five batches of 10,000 random signatures were tested. The influence of the number of genes was tested by splitting the dataset in half, this time based on genes rather than patients. Again, five batches of 10,000 random signatures were tested. To investigate the influence of the number of reporters measured per gene, again a set of five batches of 10,000 genes was tested on the dataset, considering only genes with more than one reporter. This was repeated, but for each gene only one reporter measurement was considered. Of these parameters only patient number influenced the false discovery rate. Results of the analysis to determine the influence of the number of genes and the number of reporters measured per gene are given in the supplementary data.

Influence of patient numbers

It has already been reported in previous studies^{26,27} that the number of patients influences the false discovery rate. The Miller dataset was split into two and four groups respectively to confirm the importance of this factor. The same 5 batches of 10,000 random signatures that were tested on the whole dataset were tested on

these subgroups (Figure 5.2A). Indeed the number of prognostic signatures increases dramatically when the size of the patient group decreases. To characterize the relationship between patient number and the probability that a randomly chosen signature is considered prognostic, additional analyses were performed for the batch of 10,000 runs with ten genes. The dataset was split into three, five and ten groups respectively. Figure 5.2B shows the distribution of the AUCs for the batch of 10,000 random signatures consisting of ten genes for the different dataset sizes. It is clear that the smaller the dataset, the wider and flatter the distribution becomes. Figure 5.2C presents the number of prognostic signatures as a function of dataset size.

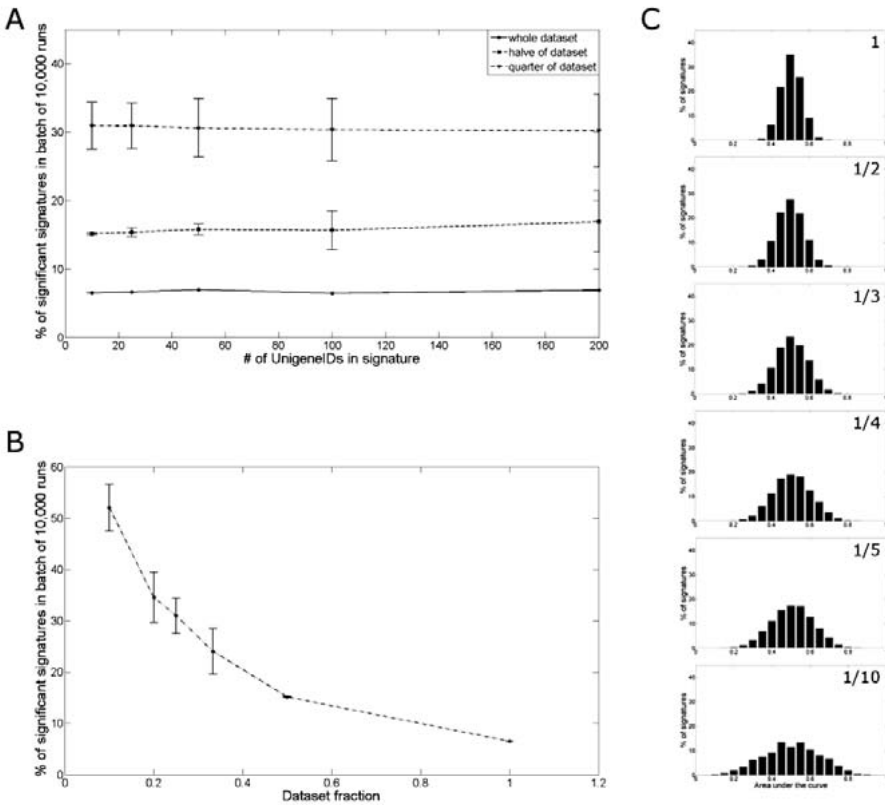


Figure 5.2

The effect of dataset size was evaluated by testing batches of random gene sets on different subsets of the Miller dataset.

Effect of filtering

One of the parameters that could account for differences in the number of prognostic signatures for a given dataset is filtering. To briefly explore the influence of filtering, two simple filtering methods were applied on the Miller dataset. After this filtering, again five batches of 10,000 signatures were tested.

The first filtering procedure was to only consider reporters that had no absent calls in the patients. This very stringent filtering resulted in a reduction in number of reporters from ~45,000 to ~7,300 (approximately 5,000 unique UnigenIDs). The second filtering method, often used in microarray based studies, consists of simply applying a threshold to the fold change. To show the effect of this step on the number of false positives a twofold threshold was applied. Only genes that show at least a two-fold change across the patients are considered. This reduced the number of reporters from ~45,000 to ~23,000.

The results for these analyses show that both filtering methods have a different effect (Figure S5.1C). Fold change filtering did not influence the probability that a randomly chosen signature is considered prognostic; rather, it provides similar results to those of non-filtered analysis. Filtering for absent reporters, on the other hand, introduced a signature size dependency for the false positive rate. A small signature size resulted in a false positive rate of ~10%, whereas large signatures had a false discovery rate of only ~0.5%. The average, however, stands at 5-6%, similar to the non-filtering and fold change filtering analyses.

Signature testing procedure

To demonstrate that this random signature method can be used with all sorts of signature evaluation methods, two additional evaluation procedures were tested in the Miller dataset. In the previous analyses the signature score was used as continuous variable.

Here we selected 10,000 random samples of 10, 25, 50, 100 and 200 genes. In the first setup the signature score was used to median dichotomize the patients. In the second setup these gene subsets were in a K-nearest neighbor classification (KNN) combined with leave-one-out-cross validation (LOOCV). Both procedures results in patient classification into two groups, which were then coupled to outcome and evaluated by the AUC. Similar AUC distributions are obtained with these different signature evaluation procedures, exact distributions characteristics differ slightly

(Table 5.4). The numbers of random gene sets passing the criteria are comparable (Supplementary Figure S5.1D).

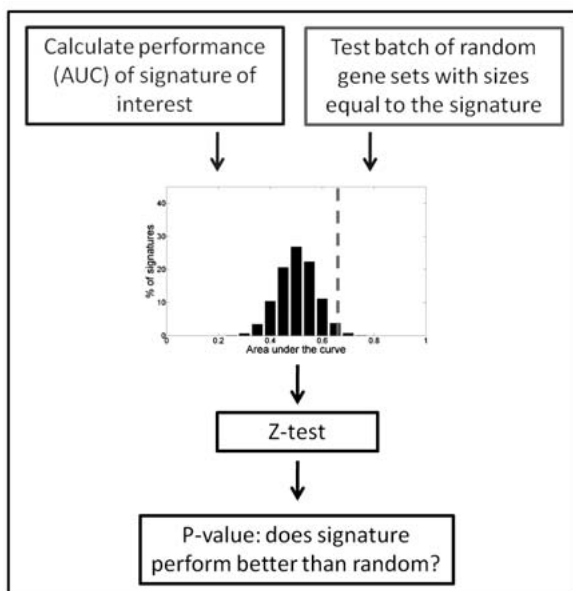


Figure 5.3

Proposed workflow for random signature testing procedure.

Table 5.4 Batches of 1,000,000 random signatures of 10, 25, 50, 100 and 200 genes were tested in the Miller dataset, where three different signature evaluation procedures were used.

| | Signature score | Median dichotomized | KNN with LOOCV |
|-------------------------|---|---|---|
| % Unigenes in signature | Average AUC (\pm standard deviation) [min - max] | Average AUC (\pm standard deviation) [min - max] | Average AUC (\pm standard deviation) [min - max] |
| 10 | 0.505 \pm 0.054 | 0.535 \pm 0.045 | 0.553 \pm 0.036 |
| | 0.312 – 0.692 | 0.348 – 0.702 | 0.426 – 0.717 |
| 25 | 0.507 \pm 0.054 | 0.535 \pm 0.045 | 0.558 \pm 0.033 |
| | 0.313 – 0.692 | 0.338 – 0.690 | 0.412 – 0.697 |
| 50 | 0.509 \pm 0.055 | 0.538 \pm 0.045 | 0.561 \pm 0.031 |
| | 0.304 – 0.694 | 0.374 – 0.671 | 0.457 – 0.678 |
| 100 | 0.513 \pm 0.053 | 0.540 \pm 0.044 | 0.562 \pm 0.030 |
| | 0.319 – 0.682 | 0.352 – 0.679 | 0.435 – 0.689 |
| 200 | 0.518 \pm 0.052 | 0.543 \pm 0.043 | 0.562 \pm 0.029 |
| | 0.335 – 0.684 | 0.373 – 0.688 | 0.449 – 0.691 |

Table 5.5 Evaluating 4 published signatures with the random signature testing procedure in three breast cancer datasets.

| Wound signature | | | | | |
|-----------------|--------------------------|--------------------|----------------------|-------|----------------------|
| Dataset | p-value log rank test | HR [95% CI] | p-value Wald test | AUC | p-value Z-test |
| Miller | 0.001 | 2.48 [1.40 – 4.41] | 0.002 | 0.671 | 0.002 |
| Wang | 0.016 | 1.59 [1.08 – 2.36] | 0.017 | 0.597 | 0.001 |
| van de Vijver | $7.27 \cdot 10^{-9}$ | 4.15 [2.42 – 7.04] | $7.71 \cdot 10^{-8}$ | 0.688 | 0.018 |
| IGS signature | | | | | |
| Dataset | p-value log rank test | HR [95% CI] | p-value Wald test | AUC | p-value Z-test |
| Miller | $5.28 \cdot 10^{-4}$ | 2.66 [1.49 – 4.78] | $7.93 \cdot 10^{-4}$ | 0.645 | 0.010 |
| Wang | $8.11 \cdot 10^{-5}$ | 2.17 [1.45 – 3.25] | $1.17 \cdot 10^{-4}$ | 0.644 | $5.80 \cdot 10^{-5}$ |
| van de Vijver | $2.16 \cdot 10^{-6}$ | 3.12 [1.88 – 5.15] | $6.35 \cdot 10^{-6}$ | 0.668 | 0.038 |
| Early 0% | | | | | |
| Dataset | p-value log rank test | HR [95% CI] | p-value Wald test | AUC | p-value Z-test |
| Miller | 0.014 | 2.00 [1.13 – 3.53] | 0.015 | 0.601 | 0.083 |
| Wang | 0.051 | 0.69 [0.47 – 1.02] | 0.055 | 0.452 | 0.001 |
| van de Vijver | 0.786 | 1.06 [0.67 – 1.67] | 0.785 | 0.508 | 0.162 |
| Early 2% | | | | | |
| Dataset | p-value log rank test | HR [95% CI] | p-value Wald test | AUC | p-value Z-test |
| Miller | 0.026 | 1.85 [1.06 – 3.22] | 0.027 | 0.593 | 0.105 |
| Wang | 0.545 | 0.89 [0.61 – 1.32] | 0.558 | 0.498 | 0.045 |
| van de Vijver | 0.321 | 1.25 [0.80 – 1.97] | 0.321 | 0.549 | 0.418 |

Evaluating signatures by random testing

To show that the random signature testing method is a valuable tool in microarray based studies, several published gene signatures were tested. In short, the suggested procedure to test a signature in a dataset is as follows (Figure 5.3). For the signature of interest the AUC was calculated in the dataset, additionally the AUC distribution for batches of random signatures with a similar size as the signature of interest was computed. The signature AUC was then compared to the random signature AUC distribution with a Z-test to assess whether the signature of interest performed better than could be expected by chance.

The Wound signature¹³, “invasiveness gene signature” (IGS)¹⁰ and two early hypoxia signatures¹⁵ are recently published gene signatures. For the Wound and IGS signatures it was previously shown that these signatures had high prognostic value in different datasets and cancer types^{10,13,28}. The two early hypoxia signatures however, were only evaluated in one dataset¹⁵. These signatures were evaluated in

the three breast cancer datasets^{8,29,30} with the signature score (details are provided in the supplementary data).

For the Miller dataset also Kaplan-Meier survival analyses were performed, since the two early hypoxia signatures were previously tested in this dataset. The results of Kaplan-Meier survival analyses and the random signature testing are given in Table 5.5 and Figure S5.3. From the Kaplan-Meier survival analyses all four signatures seemed to have a high prognostic value (p-values log-rank test <0.05). However the random signature testing procedure indicated that the two early hypoxia signatures did not perform better than chance in that dataset. Testing the four signatures in the other two breast cancer datasets indeed showed that the two early hypoxia signatures did not have prognostic value (p-values log-rank test >0.05). For the Wound and IGS signatures both evaluation procedures indicated that the performance of these signatures is high in the different datasets and that this is unlikely due to chance.

Discussion

We assessed six patient microarray datasets spanning different cancer types, numbers of patients and arrays to evaluate the effect of false positives on gene signature evaluation. Different-sized batches of 10,000 random signatures were tested in all datasets. With the given threshold, the average chance that a randomly generated signature was considered prognostic was approximately 10%, but ranged from 1% to ~40%.

Testing batches of random signatures in different datasets revealed that the AUC distribution varied widely between datasets. Choosing an arbitrary cut-off value for significance is then clearly not suited for gene signature evaluation. Rather a dataset-based cut-off value should be considered. The random testing method we propose here can be applied to calculate the level of AUC necessary to reach significance beyond random for a given signature size in a given dataset.

The random testing procedure can also be used to directly test whether the performance of a certain signature could be due to chance. A schematic overview is given in Figure 5.3. A batch of random signatures with the same size as the signature of interest can be tested along with the original signature. The AUC distribution of the random signatures can then be used to statistically test whether the original signature performs better than random. An equivalent permutation-

based validation step was used by Boutros *et al.*¹⁹ to evaluate their signature; this step provided significant information on the prognostic performance of the gene set.

We have shown that proper validation is absolutely essential in gene signature research. This supports several previous studies, which have argued that signature performance is often overestimated due to improper validation in a large number of studies^{1,26}. For several analyses, the maximum and minimum AUC were also calculated. We show that random signatures can have very high performances (AUC>0.9), which further supports this observation.

A method to overcome this multiple testing problem is validation in multiple independent datasets. We have shown that testing random signatures in two datasets decreased the chance that a random signature is called prognostic dramatically (Figure S5.2). However it is not always possible to validate a gene signature in multiple datasets. In oncology most microarray studies focus on breast and lung cancer, for these sites there are a lot of public datasets available that can be used for validation. Therefore this technique is not primarily meant for these cancer types, but rather for tumour types where only few data, in terms of the number of samples and number of datasets, are available; for those cases this technique would be valuable. By comparing the performance of four published signatures in one patient microarray dataset with Kaplan-Meier curves all signatures seemed to have prognostic value. However, applying the random testing procedure in that dataset already indicated that two out of four signatures did not perform better than chance. Testing the four signatures in multiple datasets indeed showed that these two signatures did not show prognostic value in the other datasets.

From the analyses on all six datasets, several parameters could influence the number of false positives. To assess the effect of these variables, several parameters were manipulated in one of the datasets. However, of the tested parameters, only patient number influenced the false positive rate dramatically. The need for large patient groups to obtain reliable results has already been recognised in other studies. Ntzani *et al.*²⁶ evaluated 84 microarray studies and concluded that small studies often give inflated, over-promising results. Zien *et al.*²⁷ assessed the influence of the number of samples in a different way: a simulation model was applied in which specificity and sensitivity were measured depending on changes in sample size, technical and biological variability. They showed that with small sample sizes, sensitivity and specificity were highly dependent on the biological and

technical variance, whereas larger sample sizes led to quite robust results that were less dependent on biological and technical variance. Moreover Popovici *et al.*²³ tested the effect of training set size on the performance of the trained marker in a validation dataset. Overall signature performance improved in the validation data and better concordance between training and testing results was observed when training dataset size increased.

Testing batches of random generated gene sets in different gene expression microarray datasets showed that the use of an arbitrary cut-off value for evaluation of signature significance is not suitable. Further it is important to use the same signature evaluation procedure for the random gene sets as for the signature of interest, since the AUC distribution can differ when using a different method. Thresholds should be defined for single datasets separately in order to obtain reproducible results. This permutation method can be used to establish and evaluate signature performance of any derived gene set within single or multiple datasets by comparing its performance to the performance distribution of thousands of randomly generated signatures. However it will be of most interest for cases where limited data is available.

Acknowledgements

We acknowledge financial support from Siemens Oncology Care Systems (OCS) and the CTMM framework (AIRFORCE project).

Supplementary material and methods

Data filtering and pre-processing

Datasets downloaded from the SMD^{31,32} were filtered according to the parameters in the paper. ClonIDs were chosen as gene annotation and the data obtained was log-transformed. For the normalized affymetrix arrays^{29,30} the genes were log-transformed. The Beer *et al.*³³ dataset was already pre-processed therefore to perform log-transformation all expression values below 1.1 were set to 1.1, this was similar to the processing performed by Chen *et al.*³⁴. In all other cases the data were kept in downloaded, log-transformed format⁸. ClonIDs and affymetrix probeIDs were translated into UnigenIDs (Build199) with Source (<http://smd.stanford.edu/>)

or Affymetrix data files (Affx annotation files available at www.affymetrix.com). Datasets were imported in Matlab.

AUC cut-off value definition

Ntzani *et al.*²⁶ evaluated the performance of different gene sets by means of sensitivity and specificity. These parameters were used to calculate the AUC. In their study they evaluated the performance of gene sets during cross-validation (Figure 1 upper panel in Ntzani *et al.*²⁶), on independent validation (Figure 1 lower panel in Ntzani *et al.*²⁶) and based on unsupervised approaches (Figure 2 in Ntzani *et al.*²⁶).

For each of the three figures we calculated the AUC for each data point. In Supplementary Table S5.1 the sensitivity, specificity, AUC and dataset size are provided for the three figures. From these data it is obvious that during cross-validation the performance of gene sets is high, as expected. Further it is clear that the performance of gene sets is better when datasets are small.

In Supplementary Table S5.2 the average AUC is calculated for the three figures. The use of small datasets often leads to overestimated results^{26,27}; thus, we also calculated the average AUC by omitting datasets smaller than 10, 20, 30, 40 and 50 patients respectively.

To define a threshold for AUC, not all data are included. During cross-validation the performance of a gene set is optimized and high values for AUC are found; therefore, these are omitted. Further, in our study five out of six datasets included >> 50 patients. Based on these calculations a threshold of ≤ 0.4 or ≥ 0.6 was chosen.

Evaluation of published signatures

The four published signatures evaluated here^{10,13,15} consist of gene that are both upregulated and downregulated in a certain phenotype. Therefore the definition of the score was slightly different from the score used for the random batches of signatures; a weighted average was used (Equation 2). For the Wound signature weights of -1 and 1 were assigned to genes representing a quiescent and activated wound response respectively. For the IGS signature log ratios were provided in the supplementary data of the paper, a weight of 1 is assigned to genes with a positive log ratio and a weight of -1 to genes with a negative log ratio. The two early hypoxia signatures contained only genes that are upregulated early in hypoxia, therefore all weights were 1.

$$\text{Score} = \frac{\sum_{i=1}^N w_i \cdot \text{exp}_{i,m}}{N} \quad \text{Equation (2)}$$

Where: Score, signature score; N, number of genes in the signature; w_i , weight of gene i and $\text{exp}_{i,m}$, gene expression of gene i in sample m .

Patients were subsequently grouped into a high and a low risk group for each signature by median dichotomization of the signature scores. Kaplan-Meier survival curves with log-rank tests and Cox proportional hazard ratio modelling were performed to address survival differences between the two groups.

Supplementary results

Influence of number of genes

The number of genes (UnigenelDs) might influence the probability that a randomly chosen signature is considered prognostic. To test this, the Miller dataset was used again. This time the dataset was split in two based on the genes, and five batches of 10,000 random signatures were tested on each part. Supplementary Figure 5.1A shows that the percentage of prognostic signatures is not influenced by the number of genes present on the microarray. To ascertain that the influence of the number of genes is negligible the dataset was split in ten, based on the genes. A run of 10,000 random signatures consisting of ten genes was performed. This showed that splitting the dataset into more groups based on the number of genes did not result in a change in the false positive rate. The average chance of finding a prognostic result at random was $6.9\% \pm 0.8$ for the ten groups, similar to the number found for the whole dataset.

Influence of the number of reporters measured per gene

Also the number of reporters measured per gene could influence the probability that a randomly chosen signature is considered prognostic. To test this hypothesis only the genes represented by more than one reporter on the Miller dataset were taken into account. The Miller dataset has approximately 10,000 UnigenelDs (genes) that are represented by multiple gene identifiers (probes in the microarray). First, five batches of 10,000 random signatures, consisting only of the genes mentioned, were tested. In the second run of five batches, only one reporter per gene was taken into account from this list. If the number of reporters per gene plays

a role, considering only one rather than multiple reporters was thought to increase the change that a randomly chosen signature was considered prognostic. The results given in Supplementary Figure 5.1B show that the number of reporters per gene did not influence the false positive rate.

Validation in multiple datasets

To show that the probability that a randomly chosen signature is considered prognostic will substantially decrease when random signatures are tested in two independent datasets, five batches of 10,000 random signature runs were once more performed, however now the batches of signatures were tested on two datasets in parallel. It is clear that the van de Vijver and Beer dataset had the highest number of false positives overall, therefore these datasets were analyzed in parallel. The results of the 10,000 runs on the two datasets, both separately and combined (presented in Supplementary Figure 5.2), show that the chance of finding a prognostic result at random dropped from $\geq 15\%$ to $\leq 5\%$ by combining the two datasets. Addition of another independent dataset will further decrease this number.

Supplementary figures and tables

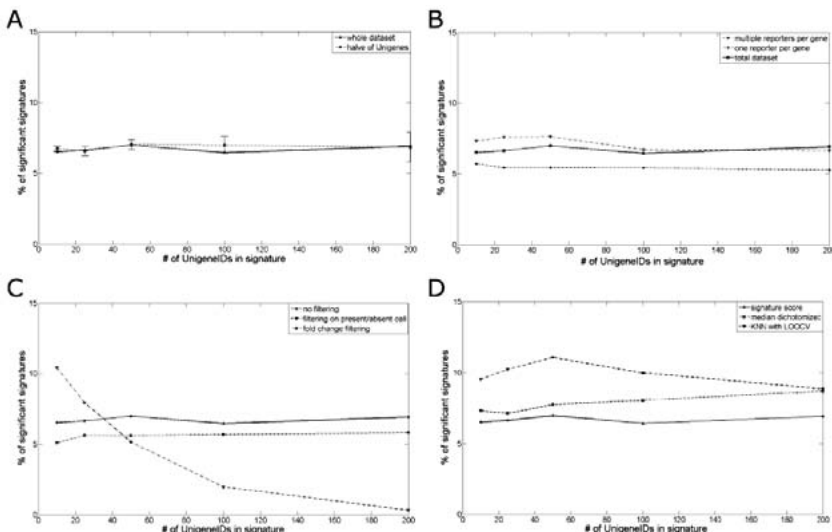


Figure S5.1

The effect of number of genes (A), number of probes per gene (B), filtering (C) and signature evaluation procedure (D) was tested in the Miller dataset.

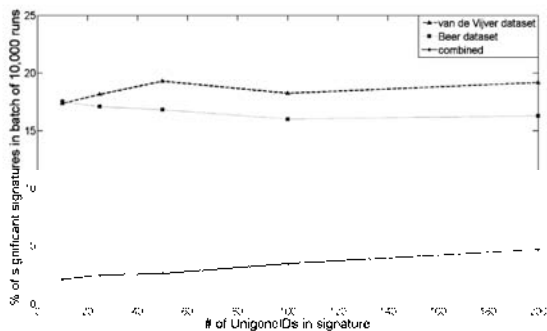


Figure S5.2

Evaluation if random gene sets in two datasets (van de Vijver and Beer dataset).

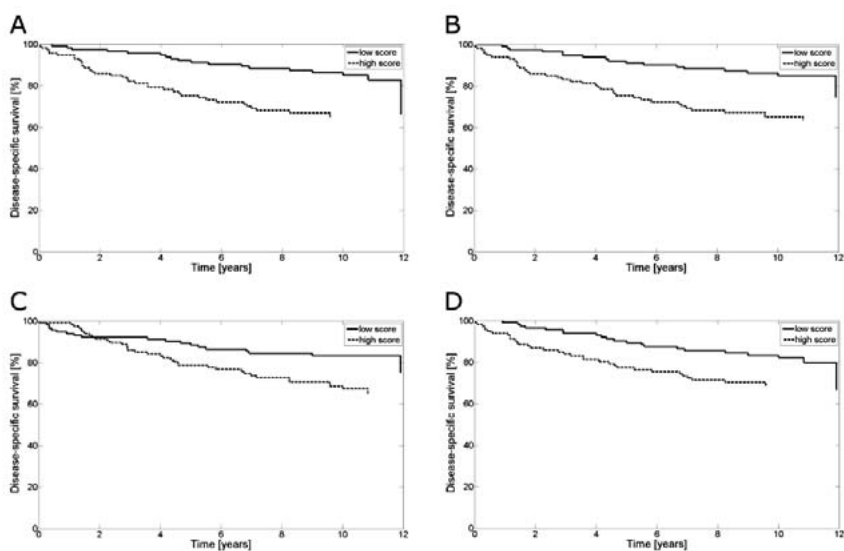


Figure S5.3

Kaplan-Meier survival analysis of 4 published gene signatures (A: Wound signature, B: IGS signature, C: early hypoxia 0% signature and D: early hypoxia 2% signature).

| Table S5.1 Calculated AUC for the gene sets evaluated in the review by Ntzani <i>et al.</i> ²⁶ | | | | |
|--|--------------------|--------------------|------------|---------------------|
| Figure 1 upper panel Ntzani <i>et al.</i>²⁶ | | | | |
| Dataset | Sensitivity | Specificity | AUC | Dataset size |
| 1 | 45 | 95 | 0.70 | 60 |
| 2 | 90 | 70 | 0.80 | 78 |
| 3 | 100 | 100 | 1 | 10 |
| 4 | 65 | 65 | 0.65 | 86 |
| 5 | 100 | 90 | 0.95 | 29 |
| 6 | 70 | 70 | 0.70 | 31 |
| 7 | 55 | 85 | 0.70 | 58 |
| 8 | 100 | 100 | 1 | 33 |
| 9 | 85 | 85 | 0.85 | 20 |
| Figure 1 lower panel Ntzani <i>et al.</i>²⁶ | | | | |
| Dataset | Sensitivity | Specificity | AUC | Dataset size |
| 1 | 100 | 70 | 0.85 | 19 |
| 2 | 90 | 50 | 0.70 | 180 |
| 3 | 60 | 65 | 0.625 | 43 |
| 4 | 45 | 65 | 0.55 | 64 |
| 5 | 40 | 55 | 0.475 | 58 |
| 6 | 70 | 65 | 0.675 | 80 |
| 7 | 85 | 90 | 0.875 | 27 |
| 8 | 75 | 100 | 0.875 | 6 |
| Figure 2 Ntzani <i>et al.</i>²⁶ | | | | |
| Dataset | Sensitivity | Specificity | AUC | Dataset size |
| 1 | 80 | 50 | 0.65 | 60 |
| 2 | 25 | 100 | 0.625 | 49 |
| 3 | 40 | 90 | 0.65 | 78 |
| 4 | 40 | 80 | 0.60 | 55 |
| 5 | 75 | 65 | 0.70 | 47 |
| 6 | 40 | 85 | 0.625 | 86 |
| 7 | 55 | 100 | 0.775 | 24 |
| 8 | 10 | 100 | 0.55 | 125 |
| 9 | 70 | 80 | 0.75 | 16 |
| 10 | 80 | 75 | 0.775 | 29 |
| 11 | 55 | 85 | 0.70 | 15 |
| 12 | 90 | 65 | 0.775 | 39 |
| 13 | 55 | 80 | 0.675 | 40 |
| 14 | 40 | 80 | 0.60 | 240 |
| 15 | 70 | 65 | 0.675 | 40 |
| 16 | 100 | 40 | 0.70 | 21 |
| 17 | 80 | 100 | 0.90 | 20 |

Table S52 Average AUC for the gene sets evaluated in the review by Ntzani *et al.*²⁶ depending on dataset size threshold.**Figure 1 upper panel** Ntzani *et al.*²⁶

| Threshold dataset size | # of datasets included | Average AUC |
|------------------------|------------------------|-------------|
| None | 9 | 0.817 |
| 10 | 8 | 0.794 |
| 20 | 7 | 0.786 |
| 30 | 6 | 0.758 |
| 40 | 4 | 0.713 |
| 50 | 4 | 0.713 |

Figure 1 lower panel Ntzani *et al.*²⁶

| Threshold dataset size | # of datasets included | Average AUC |
|------------------------|------------------------|-------------|
| None | 8 | 0.703 |
| 10 | 7 | 0.679 |
| 20 | 6 | 0.65 |
| 30 | 5 | 0.605 |
| 40 | 5 | 0.605 |
| 50 | 4 | 0.60 |

Figure 2 Ntzani *et al.*²⁶

| Threshold dataset size | # of datasets included | Average AUC |
|------------------------|------------------------|-------------|
| None | 17 | 0.690 |
| 10 | 17 | 0.690 |
| 20 | 14 | 0.670 |
| 30 | 11 | 0.648 |
| 40 | 8 | 0.625 |
| 50 | 6 | 0.613 |

References

1. Dupuy, A. & Simon, R.M. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* 99, 147-157 (2007).
2. Bryant, C.M., et al. Clinically relevant characterization of lung adenocarcinoma subtypes based on cellular pathways: an international validation study. *PLoS One* 5, e11712 (2010).
3. Fishel, I., Kaufman, A. & Ruppin, E. Meta-Analysis of Gene Expression Data: A Predictor-Based Approach. *Bioinformatics* (2007).
4. Michiels, S., Koscielny, S. & Hill, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 365, 488-492 (2005).
5. Bild, A.H., Potti, A. & Nevins, J.R. Linking oncogenic pathways with therapeutic opportunities. *Nat Rev Cancer* 6, 735-741 (2006).
6. Crijns, A.P., et al. Survival-related profile, pathways, and transcription factors in ovarian cancer. *PLoS medicine* 6, e24 (2009).
7. Miller, L.D., et al. Optimal gene expression analysis by microarrays. *Cancer Cell* 2, 353-361 (2002).
8. van de Vijver, M.J., et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347, 1999-2009 (2002).
9. Rosenwald, A., et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 346, 1937-1947 (2002).
10. Liu, R., et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *The New England journal of medicine* 356, 217-226 (2007).
11. Sanchez-Palencia, A., et al. Gene expression profiling reveals novel biomarkers in nonsmall cell lung cancer. *Int J Cancer* (2010).
12. Zhu, C.Q., et al. Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J Clin Oncol* 28, 4417-4424 (2010).
13. Chang, H.Y., et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* 2, E7 (2004).
14. Chi, J.T., et al. Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. *PLoS Med* 3, e47 (2006).
15. Seigneuric, R., et al. Impact of supervised gene signatures of early hypoxia on patient survival. *Radiother Oncol* 83, 374-382 (2007).
16. Subramanian, J. & Simon, R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Natl Cancer Inst* 102, 464-474 (2010).
17. Lau, S.K., et al. Three-gene prognostic classifier for early-stage non small-cell lung cancer. *J Clin Oncol* 25, 5562-5569 (2007).
18. Fan, C., et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 355, 560-569 (2006).
19. Boutros, P.C., et al. Prognostic gene signatures for non-small-cell lung cancer. *Proceedings of the National Academy of Sciences of the United States of America* 106, 2824-2828 (2009).
20. Jeanmougin, M., et al. Should we abandon the t-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies. *PLoS One* 5, e12336 (2010).
21. Fan, X., et al. DNA microarrays are predictive of cancer prognosis: a re-evaluation. *Clin Cancer Res* 16, 629-636 (2010).

22. Ein-Dor, L., Kela, I., Getz, G., Givol, D. & Domany, E. Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics* 21, 171-178 (2005).
23. Popovici, V., *et al.* Effect of training-sample size and classification difficulty on the accuracy of genomic predictors. *Breast Cancer Res* 12, R5 (2010).
24. Boulesteix, A.L. & Slawski, M. Stability and aggregation of ranked gene lists. *Brief Bioinform* 10, 556-568 (2009).
25. Xu, W.W. & Carter, C.J. Parallel multiplicity and error discovery rate (EDR) in microarray experiments. *BMC Bioinformatics* 11, 465 (2010).
26. Ntzani, E.E. & Ioannidis, J.P. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* 362, 1439-1444 (2003).
27. Zien, A., Fluck, J., Zimmer, R. & Lengauer, T. Microarrays: how many do you need? *J Comput Biol* 10, 653-667 (2003).
28. Chang, H.Y., *et al.* Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* 102, 3738-3743 (2005).
29. Miller, L.D., *et al.* An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A* 102, 13550-13555 (2005).
30. Wang, Y., *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365, 671-679 (2005).
31. Zhao, H., *et al.* Gene expression profiling predicts survival in conventional renal cell carcinoma. *PLoS Med* 3, e13 (2006).
32. Garber, M.E., *et al.* Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A* 98, 13784-13789 (2001).
33. Beer, D.G., *et al.* Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 8, 816-824 (2002).
34. Chen, H.Y., *et al.* A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 356, 11-20 (2007).

Chapter 6

Re: Gene expression-based prognostic signatures in lung cancer: ready for clinical use?

Published in: J Natl Cancer Inst 2010; 102(21): 1678-1679

Paul C. Boutros, Melania Pintilie, Thomas John, Maud H.W. Starmans, Sandy D. Der, Frances A. Shepherd, Ming-Sound Tsao, Igor Jurisica

Subramanian and Simon reviewed mRNA-based prognostic signatures for non-small cell lung cancer (NSCLC)¹. They concluded that many studies were poorly executed and that no gene expression signature is ready for clinical application. We agree that guidelines for design, reporting, and analysis are needed, but have reservations concerning the accuracy of their review.

We examined the 20 scores that they assigned to the two signatures reported by our group^{2,3} and found several errors. The study by Lau *et al.*³ received a score of 0 for “statistically significant improvement over standard risk factors” despite the fact that 95% confidence intervals for the difference in concordance index between a clinical model and a model containing clinical and molecular information were presented in Table A5. This table demonstrates that zero is excluded from the 95% confidence interval of the difference in a validation set from Duke (95% confidence interval 0.018 to 0.162), and was included at the far left in validation sets from Harvard (−0.018 to 0.079) and Michigan (−0.046 to 0.122). The study by Boutros *et al.*² received a score of 0 for “validation in stages IA and IB”; however, Boutros *et al.*² presented a stage-adjusted validation of 345 stage I patients in Figure S3. That study also provided a complete model specification readily replicable by experts in the field: “[Euclidean] distance between the expression profile for each patient and the cluster centers (medians).” Nevertheless, a score of 0 was assigned.

With regard to the validations that Subramanian and Simon performed, we found that our specified procedure was not followed. They tested our three-gene signature³ in an independent dataset⁴ using data preprocessed with the Model-Based Expression Indices with pseudo-count addition. Our method requires each Director’s Challenge subset to be separately pre-processed using the Robust Multiarray Average algorithm, with no pseudo-count addition. This change alters patient classification and reaffirms the need to follow author-specified procedures exactly.

We also question the appropriateness of the binary scoring criteria used by Subramanian and Simon. For example, a study that provides three of the four specified clinical covariates receives the same score as one that provides none. Of greater concern, a study that provides summary information receives the same score as one that provides individual patient annotation. Similarly, although we agree with Subramanian and Simon about the importance of adequate tissue handling, penalizing a study for not including a specific sentence detailing tissue handling presumes that such methods were not followed, when indeed they are

standard. As such, the Director's Challenge study⁴, in which a strict tissue handling protocol was enforced and described, would be given a score of 0 for omitting this specific sentence. We also disagree with the authors that studies should be penalized for showing training set performance (i.e., resubstitution statistics). Although resubstitution statistics provide an optimistically biased estimate of performance, they still have value at a number of levels. First, it is important that all results are disclosed; second, the direction of effect is important; third, the difference between the training and validation sets is a measure of over-fitting and is important for future studies; fourth, demonstration of the training set performance allows evaluation of the signature on its initial patient population. The authors view resubstitution statistics as unnecessary and undesirable. We view them as necessary, but not sufficient, for signature evaluation.

Finally, we believe that all datasets and validations should be published. Publication of failed validations provides datasets that are critical for the development and validation of more mature markers, and reduces negative publication bias⁵. Given the small sample sizes of existing datasets, validation in many independent patient populations is extremely important. According to the criteria proposed by Subramanian and Simon, only two datasets have appropriate patient characteristics, and for one of these⁶, the full data are not publicly available.

There remains a strong clinical need for improved prediction of NSCLC prognosis⁷. A systematic review is timely, but the authors of such reviews — especially when scoring studies according to inclusion of specific data — have a responsibility to present complete and accurate results.

References

1. Subramanian, J. & Simon, R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Natl Cancer Inst* **102**, 464-474 (2010).
2. Boutros, P.C., *et al.* Prognostic gene signatures for non-small-cell lung cancer. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2824-2828 (2009).
3. Lau, S.K., *et al.* Three-gene prognostic classifier for early-stage non small-cell lung cancer. *J Clin Oncol* **25**, 5562-5569 (2007).
4. Shedden, K., *et al.* Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* **14**, 822-827 (2008).
5. Kyzas, P.A., Denaxa-Kyza, D. & Ioannidis, J.P. Almost all articles on cancer prognostic markers report statistically significant results. *Eur J Cancer* **43**, 2559-2579 (2007).
6. Roepman, P., *et al.* An immune response enriched 72-gene prognostic profile for early-stage non-small-cell lung cancer. *Clin Cancer Res* **15**, 284-290 (2009).
7. Butts, C.A., *et al.* Randomized phase III trial of vinorelbine plus cisplatin compared with observation in completely resected stage IB and II non-small-cell lung cancer: updated survival analysis of JBR-10. *J Clin Oncol* **28**, 29-34 (2010).

Chapter 7

Validation of a microarray-based prognostic marker for non-small cell lung cancer: sensitivity to data pre-processing

Submitted

Maud H.W. Starmans, Melania Pintilie, Thomas John, Sandy D. Der, Frances A. Shepherd, Igor Jurisica, Philippe Lambin, Ming-Sound Tsao, Paul C. Boutros

Abstract

Independent validation is a critical step in the development and clinical application of a new biomarker. We tested two of our previously published prognostic multi-gene markers for non-small cell lung cancer (NSCLC) in an independent 442-patient dataset. Both markers validated successfully in the entire cohort (3-gene marker: HR = 1.63, $P = 0.001$, 6-gene marker: HR = 1.42, $P = 0.030$). Further, despite being underpowered for a stage-specific analysis, the 3-gene marker significantly prognosed both Stage IB and Stage II patients. A recent report failed to replicate these results in the same independent dataset, so we systematically investigated reasons for this discrepancy. Differences in microarray pre-processing methodologies were identified as a key factor. By examining 24 separate pre-processing techniques we show that minor alterations in pre-processing can change a successful prognostic marker (HR = 1.85; $P < 0.001$) into one indistinguishable from random chance (HR = 1.15; $P = 0.348$). We replicated this sensitivity to data pre-processing in a second dataset. Continued validation of prognostic multi-gene markers in independent and larger datasets is essential; proper validation must follow the pre-processing steps of the original study.

Despite improvements in staging, surgical methodologies, chemotherapy regimens and the addition of adjuvant therapies, 30-50% of patients with resectable non-small cell lung cancer (NSCLC) suffer relapse and die within 5 years¹⁻³. As more molecular-targeted therapies are developed, there is a clear and emerging clinical need for additional prognostic factors to sub-classify patients better in order to guide personalized therapy.

Several groups, including our own, have employed transcriptome-profiling of surgically excised tumor samples to develop multi-gene prognostic markers (sometimes called prognostic signatures or classifiers)⁴⁻⁷. However, there is minimal gene-wise overlap between these multi-gene markers⁸, and challenges exist in the datasets and analyses used to generate them⁹. The Director's Challenge NSCLC study was an attempt to provide a large, sufficiently-powered dataset to discover reproducible multi-gene markers¹⁰. This study integrated four independent datasets of adenocarcinomas named according to the institutions at which they were generated: UM, HLM, MSK and CAN/DF. None of the multi-gene markers tested in that study were validated for the primary end-point of stage I survival.

We previously developed two multi-gene markers derived from PCR-analysis of a selected group of candidate prognostic markers^{8,11}. While independent validations were described in the original reports, it is critical that a good biomarker works in additional large patient cohorts. Therefore, we used the Director's Challenge dataset to interrogate the validity of our two multi-gene markers. We followed the exact procedures described in the original studies (see Supplementary Material) and evaluated our markers using stage-adjusted Cox proportional hazards models. Both our 3-gene marker (Figure 7.1A; HR = 1.63, $P = 1.18 \times 10^{-3}$, Wald test) and 6-gene marker (Figure 7.1B; HR = 1.42, $P = 3.01 \times 10^{-2}$, Wald test) successfully validated.

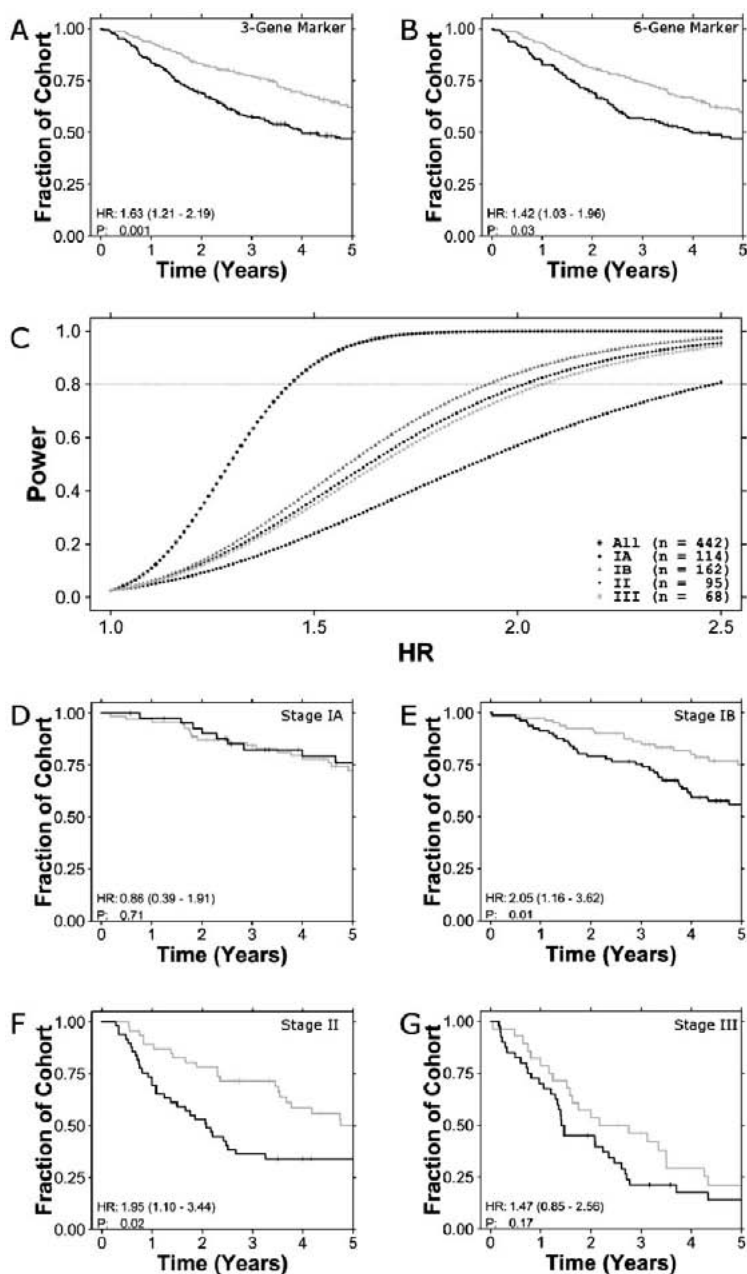


Figure 7.1

The previously published 3-gene⁸ (A) and 6-gene¹¹ (B) markers were validated in the DC dataset¹⁰. Power calculations were performed (C) prior to a sub-stage analysis for the 3-gene marker (D: stage IA, E: stage IB, F: stage II and G: stage III). In gray a threshold line is drawn for power of 0.8 (C). [HR: hazard ratio, P: p-value Wald test].

Recently, Subramanian and Simon⁹ performed a critical review of a number of prognostic multi-gene markers for NSCLC. As part of this review, they suggested analyzing patients of specific stages independently and applied this approach to the Director's Challenge dataset. Before embarking on a similar analysis we performed a power analysis to estimate the likelihood that real differences could be identified in each subgroup. Figure 7.1C shows that even in the large Director's Challenge cohort, these sub-group analyses are marginally powered; there is only a 57% chance of detecting a real hazard ratio of 2.0 in stage IA patients. It is therefore not surprising that both our markers were ineffective at classifying stage IA patients (3-gene marker: Figure 7.1D, HR = 0.86, $P = 0.710$, Wald test, 6-gene marker: HR = 0.69, $P = 0.424$ Wald test). Nevertheless, the 3-gene marker was validated in stage IB (Figure 7.1E, HR = 2.05, $P = 1.41 \times 10^{-2}$, Wald test) and stage II patients (Figure 7.1F, HR = 1.95, $P = 2.11 \times 10^{-2}$, Wald test). The 6-gene marker was validated in stage II patients (HR = 1.90, $P = 3.50 \times 10^{-2}$). These results emphasize the importance of continued validation on new datasets, as even the largest existing cohorts are insufficiently powered.

As part of their review, Subramanian and Simon⁹ attempted to validate this same 3-gene marker on the Director's Challenge dataset. Surprisingly their results differed from those described above. They observed no validation for stage IB ($P = 0.35$), leading them to suggest that current multi-gene markers lack clinical utility. We investigated the origins of this discrepancy and identified differences in microarray data pre-processing. As described in our original report⁸, we used RMA-normalized data with each site in the Director's Challenge treated separately. Subramanian and Simon⁹ used MBEI quantile-normalized data with pseudo-count addition, while treating all sites together. This issue has been appreciated in other studies where differences in pre-processing contributed to failed reproduction of drug sensitivity predictions^{12,13}.

To better understand the effect of different pre-processing strategies, we analyzed the Director's Challenge dataset using a panel of different methods and tested our 3-gene prognostic marker against each. We investigated four separate factors. First, we compared treating the Director's Challenge study as a single study or as four site-specific datasets (the original report indicated high inter-site variability¹⁰). Second, we employed four diverse and commonly-used pre-processing algorithms¹⁴⁻¹⁸. Third, we evaluated the effects of \log_2 -transformation, a standard operation in microarray analysis. Finally, both default Affymetrix gene annotations and updated

Entrez Gene-based annotations were tested¹⁹. All combinations of 2 dataset handling, 6 pre-processing and 2 annotation methods were evaluated, for a total of 24 datasets. We evaluated the performance of the 3-gene prognostic marker on each dataset in both overall and stage-specific models (see Supplementary Material). Figure 7.2 outlines this procedure; Supplementary Data File 1 gives the classification of every patient in each dataset, and all technical details are presented in Supplementary Material.

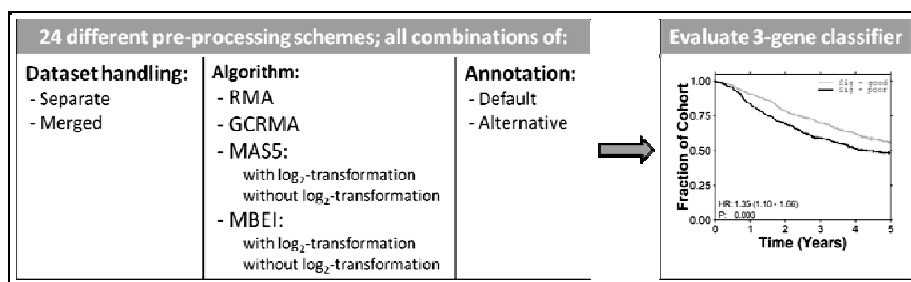


Figure 7.2

Scheme to test marker performance sensitivity to different pre-processing schedules. The DC dataset¹⁰ will be pre-processed with 24 different schemes. Subsequently, the 3-gene marker⁸ is evaluated.

Our systematic analysis demonstrates that the validation of multi-gene markers is highly sensitive to data pre-processing (Supplementary Table S7.1). This is especially true in stage-specific analyses: hazard ratios for Stage IB patients range from 0.89 ($P = 0.672$, Wald test) to 2.05 ($P = 1.41 \times 10^{-2}$, Wald test). Small changes in pre-processing led to large changes in classification performance. For example, changing the pre-processing algorithm from RMA to MBEI changed validation in Stage IB patients from success to failure (compare Figures 7.1E and 7.3A). Similarly in the total patient group reduced prognostic power was observed (Figure 7.1A; HR = 1.63, $P = 1.18 \times 10^{-3}$, Wald test vs. Figure 7.3B: HR = 1.36, $P = 3.88 \times 10^{-2}$, Wald test). Performance within a single method varies across stage (Figure 7.3C).

These differences in classifier performance are caused by changes in the classification status of a significant portion of patients. Only 151 out of 442 patients are classified identically by all 24 pre-processed schemes (Figure 7.3D). About equal numbers of unchanged classifications are in the good (77) and poor (74) prognosis groups.

To validate that patient classification is highly affected by differences in pre-processing the same procedure (Figure 7.2) was applied to the Bild dataset²⁰. Again, only 45 out of 111 patients are classified identically across the 24 pre-processing methodologies (Figure 7.2E, Supplementary Data File 2) resulting in large differences in 3-gene marker success (Figure 7.3F, 7.3G and Supplementary Table S7.2). Interestingly, 3-gene marker performance improved significantly in both datasets when only patients identically classified with all 24 pre-processing schedules were considered (Figure 7.3H and 7.3I).

Our findings concord with previous observations that pre-processing procedures influence downstream analysis of differentially regulated genes and gene networks^{14,21,22}. It has been reported that the optimal pre-processing method is intrinsically dependent on experimental design and cannot be standardized^{22,23}. The pre-processing scheme used for original validation of the 3-gene marker was again one of the best-performing in this second validation (Figure 7.3C).

In summary, we report the successful validation of our 3- and 6-gene prognostic markers for NSCLC⁸ in the 442-patient Director's Challenge dataset. Despite being underpowered, the marker significantly prognosed stage IB and stage II patients. In the course of this validation we discovered an extreme sensitivity to the pre-processing methodology. It was recently stated that "The differences in the preprocessing steps for microarray data are immaterial... [when] the original classifier was developed using RT-qPCR."²⁴ Our results demonstrate that this statement is incorrect. Instead, they indicate that different pre-processing methodologies can be used to identify and validate different multi-gene markers from the same dataset. Continued validation of prognostic multi-gene markers in newer and larger datasets is important, but for emerging genomic technologies it will be critical to exactly follow a consistent and clearly-reported data pre-processing schedule to accurately gauge clinical applicability.

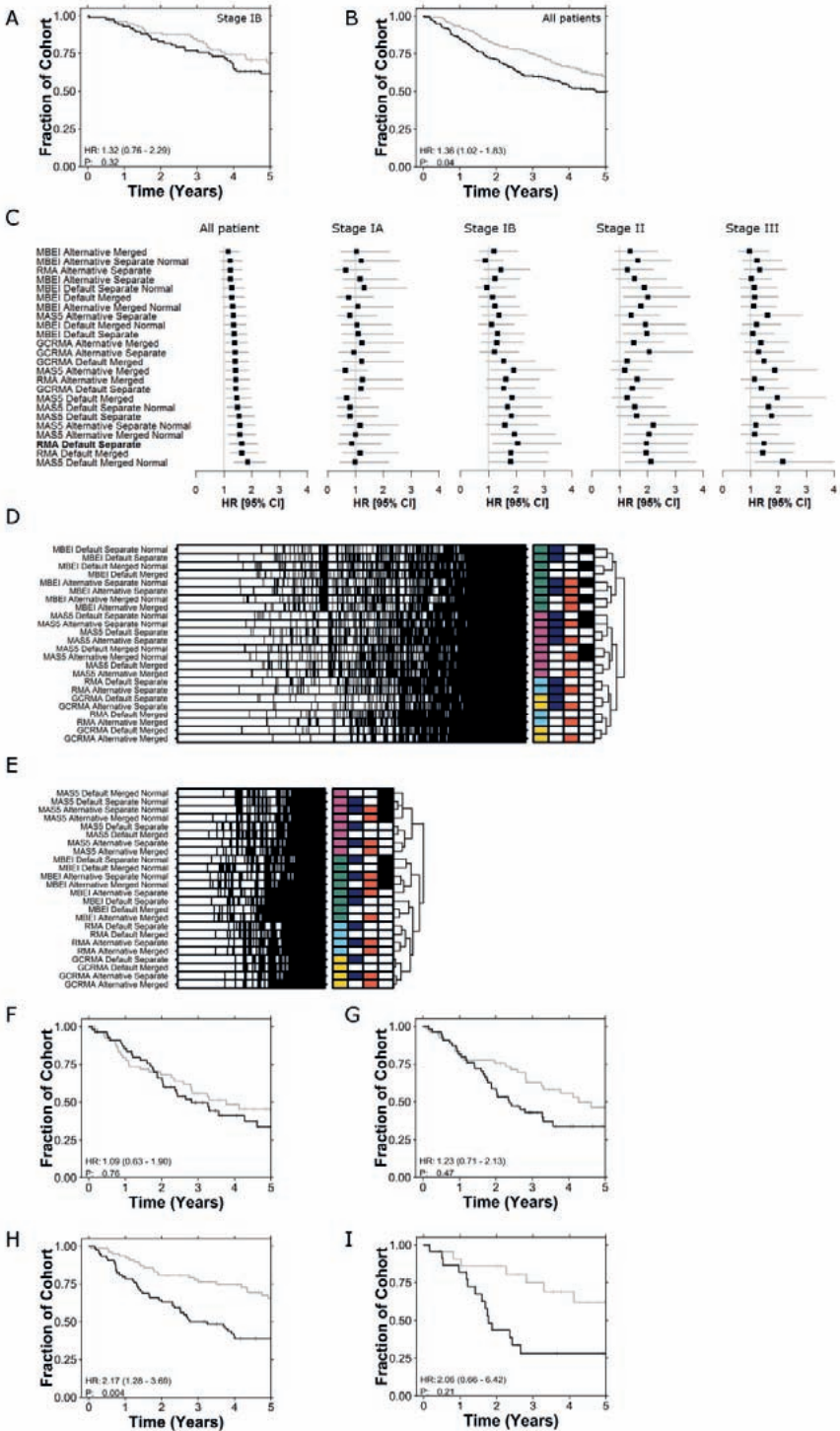


Figure 7.3 (previous page)

Pre-processing influences marker validation: Changing the pre-processing algorithm (to the algorithm used by Subramanian and Simon⁹) resulted in failed validation of the 3-gene marker⁸ in the stage IB patients (A) and total cohort (B). Results for all these analyses are summarized in Forest plots (C). [Boxes : HR, lines: 95% confidence interval]. Classifications in the 24 different schedules are visualized in a heatmap (D). [White indicates a patient predicted to have good prognosis and black indicates a patient predicted to have poor prognosis. Colored sidebar displays the different pre-processing schemes, from left to right: algorithm (Dark grey: MBEI, White: MAS5, Light grey: RMA, Black: GCRMA), dataset handling (Dark grey: separate, White: merged), annotation method (White: default, Light grey: alternative), and log₂-transformed or not (White: yes, Black: no)].

This was also tested in a second dataset²⁰ for the 3-gene marker (E: heatmap, F: MAS5, separate dataset handling, default annotation; G: RMA, separate dataset handling, default annotation). Including only patients with identical classifications across all pre-processing schemes improved 3-gene marker performance in both datasets (H and I). [HR: hazard ratio, P: p-value Wald test].

Acknowledgements

This work was supported by the Center for Translational Molecular Medicine (www.ctmm.nl) (AIRFORCE Project Ref. 030-103) and the Ontario Institute for Cancer Research through funding provided by the Government of Ontario to PCB.

Supplementary material and methods

Classifier evaluation

All analyses were performed in R (version 2.11.1). The Director's Challenge dataset¹⁰ was used to validate our previously published 3-gene and 6-gene markers^{8,11}. Data processing and patient classification were performed as in the original studies^{8,11}. The data consists of four independent datasets; UM, HLM, MSK and CAN/DF. Shedden *et al*¹⁰ reported high inter-group variability in their study; therefore these datasets were treated separately. Further pre-processing was performed with RMA¹⁴ (R package: affy version 1.28.0).

On a microarray every gene is represented by a set of oligonucleotides, called a ProbeSet. The annotation of these ProbeSets has changed significantly over time so

Dai *et al.*¹⁹ developed a method to generate updated ProbeSet annotations, which can be used as alternatives to standard annotation files provided. ProbeSet annotation was done with Affymetrix provided annotation (R packages: `hgu133aprobe_2.6.0` and `hgu133acdf_2.6.0`), which is referred to as default. For marker evaluation the corresponding Affymetrix ProbeSets from the original study were used (Supplementary Table S7.3). Once these genes were selected, median scaling and housekeeping gene normalization (to the geometric mean of ACTB, BAT1, B2M and TBP levels) was subsequently performed, exactly as described previously²⁵.

The 3-gene classifier involves CCR7, HIF1A and STX1A. These genes were subjected to statistical scaling and were then median-dichotomized, as outlined in Lau *et al.*⁸ The risk score was then calculated from the scaled, normalized expression as:

$$\text{Score} = 4 \times \text{STX1A}_{\text{expr}} + 3 \times \text{HIF1A}_{\text{expr}} - 3 \times \text{CCR7}_{\text{expr}}$$

In this equation $\text{STX1A}_{\text{expr}}$ for a patient is one if the patient has a level of STX1A (after applying all pre-processing schemes described above) above the median for all patients in the dataset and zero otherwise. The values of $\text{HIF1A}_{\text{expr}}$ and $\text{CCR7}_{\text{expr}}$ are calculated in the same way. Patients were then classified into risk groups based on this score. Patients with a score ≤ 2 were considered to have a good prognosis, whereas patients scoring more than 2 were classified as having a poor prognosis.

For the 6-gene classifier, Euclidean distances to the training cluster centers (median) from the original study were calculated to classify each patient¹¹. A patient was classified into the nearer cluster if the ratio of the distances between the profile and the two clusters was at least 0.9.

Prognostic performance of the 3-gene and 6-gene classifier was evaluated by stage-adjusted Cox proportional hazard ratio modeling followed by the Wald test. Five-year survival was used as endpoint. The analysis was performed using the survival library in R (version 2.35-8). A power analysis was performed to estimate the likeliness that real differences could be identified in stage-dependent patient subgroups (R package: `Hmisc_3.8-2`).

Dataset pre-processing

To assess the influence of different pre-processing schedules on signature performance the Director's Challenge dataset was pre-processed in 24 different ways. Subsequently our 3-gene classifier was evaluated in these differently pre-processed datasets.

First pre-processing for the four datasets was either done for all datasets separately or for all combined. When datasets were treated separately pre-processing was performed separately for each of the four datasets. Then patients predicted as having good prognosis in any of the datasets were pooled into one group and patients predicted as having poor prognosis in any of the datasets were pooled into another group. Then four different pre-processing algorithms were applied; RMA¹⁴, GCRMA¹⁵, MAS5^{16,17} and MBEI¹⁸ (R packages: affy version 1.26.1, gcrma version 2.20.0). In Supplementary Table S7.4 a brief description of the algorithms is given. RMA and GCRMA provide data in \log_2 -transformed space, whereas MAS5 and MBEI provide data in normal space. It is common to \log_2 -transform MAS5 and MBEI pre-processed data, and we therefore tested them both in standard and \log_2 -transformed space. Further ProbeSet annotation was done with default annotation (R packages: hgu133aprobe_2.6.0 and hgu133acdf_2.6.0), or with updated Entrez Gene-based annotations¹⁹, referred to as alternative (R packages: hgu133ahsentrezprobe_12.1.0 and hgu133ahsentrezgcdf_12.1.0). All possible combinations of pre-processing (6), annotation (2) and dataset handling (2) were made, resulting in 24 different sets. When the default Affymetrix gene-annotation was applied, the corresponding Affymetrix ProbeSets from the original study were used. When the alternative Entrez Gene ID ProbeSet annotation was utilized, matching was performed based on Entrez Gene ID. Supplementary Table S7.3 below lists the specific ProbeSets used for each gene according to each annotation protocol. Supplementary Data file 1 gives the key clinical data for each patient, along with the good/poor classifications for the 3-gene classifier in each of the pre-processing methods.

A similar procedure was applied to the Bild dataset²⁰ (R packages: affy version 1.28.0, gcrma version 2.22.0). This dataset consists of 2 batches; therefore the same 24 pre-processing schedules were applied. Default and alternative ProbeSet annotation were performed with the appropriate R packages (default: hgu133plus2probe_2.7.0, hgu133plus2cdf_2.7.0 and alternative: hgu133plus2hsentrezprobe_13.0.0, hgu133plus2hsentrezgcdf_13.0.0). The specific ProbeSets used for each gene according to each annotation protocol are listed in Table S7.3. Supplementary Data file 2 gives the key clinical data for each patient, along with the good/poor classifications for the 3-gene classifier in each of the pre-processing methods.

Supplementary tables

Table S7.1 Results of (stage-adjusted) Cox proportional hazard ratio modeling 3-gene classifier for all and stage IB patients in the 24 different pre-processed DC datasets. Significant results ($p < 0.05$) are given in bold. [HR: hazard ratio, P: p-value Wald test].

| Algorithm | Pre-processing | | | All patients | | Stage IB patients | |
|-----------|----------------|-------------|-----------|------------------------------|---|------------------------------|---|
| | Dataset | Annotation | Transform | HR [95% CI] | P | HR [95% CI] | P |
| RMA | Merged | Alternative | \log_2 | 1.42 [1.06 – 1.92] | 2.07×10^{-2} | 1.62 [0.93 - 2.82] | 8.87×10^{-2} |
| RMA | Merged | Default | \log_2 | 1.64 [1.22 – 2.21] | 1.11×10^{-3} | 1.80 [1.03 - 3.14] | 3.97×10^{-2} |
| RMA | Separate | Alternative | \log_2 | 1.23 [0.92 – 1.64] | 1.72×10^{-1} | 1.44 [0.83 - 2.49] | 1.90×10^{-1} |
| RMA | Separate | Default | \log_2 | 1.63 [1.21 – 2.19] | 1.18×10^{-3} | 2.05 [1.16 - 3.62] | 1.41×10^{-2} |
| GCRMA | Merged | Alternative | \log_2 | 1.38 [1.03 – 1.85] | 2.88×10^{-2} | 1.29 [0.75 - 2.23] | 3.55×10^{-1} |
| GCRMA | Merged | Default | \log_2 | 1.40 [1.05 – 1.88] | 2.38×10^{-2} | 1.54 [0.89 - 2.66] | 1.20×10^{-1} |
| GCRMA | Separate | Alternative | \log_2 | 1.39 [1.04 – 1.87] | 2.76×10^{-2} | 1.21 [0.70 - 2.09] | 5.01×10^{-1} |
| GCRMA | Separate | Default | \log_2 | 1.43 [1.06 – 1.91] | 1.78×10^{-2} | 1.54 [0.90 - 2.66] | 1.18×10^{-1} |
| MAS5 | Merged | Alternative | \log_2 | 1.41 [1.05 – 1.90] | 2.33×10^{-2} | 1.90 [1.06 - 3.39] | 3.01×10^{-2} |
| MAS5 | Merged | Default | \log_2 | 1.46 [1.08 – 1.97] | 1.42×10^{-2} | 1.84 [1.04 - 3.26] | 3.65×10^{-2} |
| MAS5 | Separate | Alternative | \log_2 | 1.34 [0.99 – 1.80] | 5.18×10^{-2} | 1.37 [0.78 - 2.40] | 2.69×10^{-1} |
| MAS5 | Separate | Default | \log_2 | 1.55 [1.15 – 2.08] | 4.10×10^{-3} | 1.82 [1.03 - 3.19] | 3.85×10^{-2} |
| MAS5 | Merged | Alternative | None | 1.57 [1.17 – 2.12] | 3.07×10^{-3} | 1.93 [1.10 - 3.39] | 2.27×10^{-2} |
| MAS5 | Merged | Default | None | 1.85 [1.37 – 2.50] | 6.01×10^{-5} | 1.79 [1.02 - 3.12] | 4.20×10^{-2} |
| MAS5 | Separate | Alternative | None | 1.57 [1.17 – 2.11] | 3.02×10^{-3} | 1.59 [0.91 - 2.77] | 1.01×10^{-1} |
| MAS5 | Separate | Default | None | 1.49 [1.11 – 1.99] | 7.79×10^{-3} | 1.68 [0.97 - 2.92] | 6.66×10^{-2} |
| MBEI | Merged | Alternative | \log_2 | 1.15 [0.86 – 1.54] | 3.48×10^{-1} | 1.19 [0.69 - 2.06] | 5.37×10^{-1} |
| MBEI | Merged | Default | \log_2 | 1.28 [0.96 – 1.72] | 9.70×10^{-2} | 1.14 [0.66 - 1.97] | 6.45×10^{-1} |
| MBEI | Separate | Alternative | \log_2 | 1.23 [0.92 – 1.65] | 1.69×10^{-1} | 1.23 [0.71 - 2.13] | 4.62×10^{-1} |
| MBEI | Separate | Default | \log_2 | 1.36 [1.02 – 1.83] | 3.88×10^{-2} | 1.32 [0.76 - 2.29] | 3.23×10^{-1} |
| MBEI | Merged | Alternative | None | 1.32 [0.98 – 1.76] | 6.58×10^{-2} | 1.23 [0.71 - 2.11] | 4.66×10^{-1} |
| MBEI | Merged | Default | None | 1.34 [1.00 – 1.80] | 4.74×10^{-2} | 1.10 [0.64 - 1.90] | 7.25×10^{-1} |
| MBEI | Separate | Alternative | None | 1.22 [0.91 – 1.63] | 1.82×10^{-1} | 0.89 [0.52 - 1.53] | 6.72×10^{-1} |
| MBEI | Separate | Default | None | 1.28 [0.95 – 1.72] | 1.02×10^{-1} | 0.94 [0.54 - 1.62] | 8.20×10^{-1} |

Table S7.2 Results (stage-adjusted) Cox proportional hazard ratio modeling 3-gene classifier for all patients in the 24 different pre-processed Bild datasets. Significant results ($p < 0.05$) are given in bold. [HR: hazard ratio, P: p-value Wald test].

| Algorithm | Pre-processing | | | All patients | |
|-----------|----------------|-------------|-----------|-----------------------|-----------------------|
| | Dataset | Annotation | Transform | HR [95% CI] | P |
| RMA | Merged | Alternative | \log_2 | 1.54 [0.83 – 2.86] | 1.69×10^{-1} |
| RMA | Merged | Default | \log_2 | 1.02 [0.58 – 1.79] | 9.43×10^{-1} |
| RMA | Separate | Alternative | \log_2 | 1.08 [0.60 – 1.94] | 7.97×10^{-1} |
| RMA | Separate | Default | \log_2 | 1.23 [0.71 – 2.13] | 4.66×10^{-1} |
| GCRMA | Merged | Alternative | \log_2 | 1.06 [0.58 – 1.94] | 8.43×10^{-1} |
| GCRMA | Merged | Default | \log_2 | 0.99 [0.55 – 1.77] | 9.68×10^{-1} |
| GCRMA | Separate | Alternative | \log_2 | 0.94 [0.51 – 1.73] | 8.44×10^{-1} |
| GCRMA | Separate | Default | \log_2 | 0.99 [0.55 – 1.77] | 9.74×10^{-1} |
| MAS5 | Merged | Alternative | \log_2 | 0.96 [0.52 – 1.76] | 8.87×10^{-1} |
| MAS5 | Merged | Default | \log_2 | 1.00 [0.57 – 1.76] | 9.89×10^{-1} |
| MAS5 | Separate | Alternative | \log_2 | 1.10 [0.64 – 1.90] | 7.25×10^{-1} |
| MAS5 | Separate | Default | \log_2 | 1.09 [0.63 – 1.90] | 7.56×10^{-1} |
| MAS5 | Merged | Alternative | None | 1.64 [0.94 – 2.86] | 8.48×10^{-2} |
| MAS5 | Merged | Default | None | 1.65 [0.90 – 3.03] | 1.03×10^{-1} |
| MAS5 | Separate | Alternative | None | 1.54 [0.89 – 2.69] | 1.25×10^{-1} |
| MAS5 | Separate | Default | None | 1.36 [0.77 – 2.40] | 2.95×10^{-1} |
| MBEI | Merged | Alternative | \log_2 | 1.10 [0.60 – 2.00] | 7.59×10^{-1} |
| MBEI | Merged | Default | \log_2 | 1.05 [0.57 – 1.94] | 8.76×10^{-1} |
| MBEI | Separate | Alternative | \log_2 | 1.07 [0.60 – 1.92] | 8.16×10^{-1} |
| MBEI | Separate | Default | \log_2 | 0.85 [0.46 – 1.57] | 6.06×10^{-1} |
| MBEI | Merged | Alternative | None | 1.13 [0.62 – 2.04] | 6.91×10^{-1} |
| MBEI | Merged | Default | None | 2.15 [1.15 – 4.03] | 1.67×10^{-2} |
| MBEI | Separate | Alternative | None | 1.02 [0.57 – 1.84] | 9.44×10^{-1} |
| MBEI | Separate | Default | None | 1.09 [0.61 – 1.95] | 7.79×10^{-1} |

References

1. Goldstraw, P., *et al.* The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. *J Thorac Oncol* **2**, 706-714 (2007).
2. Pignon, J.P., *et al.* Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group. *J Clin Oncol* **26**, 3552-3559 (2008).
3. Jemal, A., *et al.* Annual report to the nation on the status of cancer, 1975-2005, featuring trends in lung cancer, tobacco use, and tobacco control. *J Natl Cancer Inst* **100**, 1672-1694 (2008).
4. Beer, D.G., *et al.* Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature medicine* **8**, 816-824 (2002).
5. Bhattacharjee, A., *et al.* Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 13790-13795 (2001).
6. Larsen, J.E., *et al.* Gene expression signature predicts recurrence in lung adenocarcinoma. *Clin Cancer Res* **13**, 2946-2954 (2007).
7. Raponi, M., *et al.* Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer research* **66**, 7466-7472 (2006).
8. Lau, S.K., *et al.* Three-gene prognostic classifier for early-stage non small-cell lung cancer. *J Clin Oncol* **25**, 5562-5569 (2007).
9. Subramanian, J. & Simon, R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Natl Cancer Inst* **102**, 464-474 (2010).
10. Shedden, K., *et al.* Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nature medicine* **14**, 822-827 (2008).
11. Boutros, P.C., *et al.* Prognostic gene signatures for non-small-cell lung cancer. *Proc Natl Acad Sci U S A* **106**, 2824-2828 (2009).
12. Coombes, K.R., Wang, J. & Baggerly, K.A. Microarrays: retracing steps. *Nature medicine* **13**, 1276-1277; author reply 1277-1278 (2007).
13. Baggerly, K.A., Coombes, K.R. & Neeley, E.S. Run batch effects potentially compromise the usefulness of genomic signatures for ovarian cancer. *J Clin Oncol* **26**, 1186-1187; author reply 1187-1188 (2008).
14. Irizarry, R.A., *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-264 (2003).
15. Wu, Z., Irizarry, R.A., Gentleman, R., Martinez-Murillo, F. & Spencer, F. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. *Journal of the American Statistical Association* **99**, 909-917 (2004).
16. Liu, W.M., *et al.* Analysis of high density expression microarrays with signed-rank call algorithms. *Bioinformatics* **18**, 1593-1599 (2002).
17. Hubbell, E., Liu, W.M. & Mei, R. Robust estimators for expression analysis. *Bioinformatics* **18**, 1585-1592 (2002).
18. Li, C. & Wong, W.H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 31-36 (2001).

19. Dai, M., *et al.* Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res* **33**, e175 (2005).
20. Bild, A.H., Potti, A. & Nevins, J.R. Linking oncogenic pathways with therapeutic opportunities. *Nature reviews* **6**, 735-741 (2006).
21. Lim, W.K., Wang, K., Lefebvre, C. & Califano, A. Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks. *Bioinformatics* **23**, i282-288 (2007).
22. Hoffmann, R., Seidl, T. & Dugas, M. Profound effect of normalization on detection of differentially expressed genes in oligonucleotide microarray data analysis. *Genome Biol* **3**, RESEARCH0033 (2002).
23. Schmid, R., *et al.* Comparison of normalization methods for Illumina BeadChip(R) HumanHT-12 v3. *BMC Genomics* **11**, 349 (2010).
24. Subramanian, J. & Simon, R. Re: Gene Expression-Based Prognostic Signatures in Lung Cancer: Ready for Clinical Use? - Author Reply. *J Natl Cancer Inst* **102**, 1678-1679 (2010).
25. Barsyte-Lovejoy, D., *et al.* The c-Myc oncogene directly induces the H19 noncoding RNA by allele-specific binding to potentiate tumorigenesis. *Cancer research* **66**, 5330-5337 (2006).

Chapter 8

Discussion and future perspectives

Discussion

Personalized medicine aims to deliver, to each patient, a treatment tailored towards their specific disease and genetics in order to achieve optimal treatment outcome with minimal side effects. This requires tools that can be used to select the most appropriate therapy regimen for specific patients. A wide variety of approaches from a range of research disciplines have been applied to identify prognostic and predictive markers to improve accuracy of treatment selection. In particular, approaches based on imaging^{1,2}, serum markers^{3,4} and gene-expression signatures⁵⁻⁸ have been identified that could be used to select patient subgroups for certain therapies. Advances in these research fields are creating opportunities to slowly move away from the one-size fits all strategy. Currently tumor staging is predominantly used to make treatment decisions for most cancer types. The work presented in this thesis focuses on the application of transcriptomic biomarkers for prognostic purposes and methods to improve their clinical utility.

The role of proliferation in prognostic signatures

Since the introduction of techniques for high-throughput gene-expression profiling, numerous gene-expression signatures have been identified that showed promising results in the original studies⁹⁻¹¹. However minimal gene-wise overlap, disappointing validation in external datasets, and lack of consistency have lead to significant skepticism regarding gene expression-based classifiers¹²⁻¹⁶. On the other hand increasing data suggests that, although overlap in signatures is small, there is high consistency in their classification results. Multiple studies, including several meta-analyses, show that prognostic gene expression markers assign a large subset of patients to the same prognostic groups¹⁷⁻²¹. These data support the hypothesis that multiple gene expression-based markers track similar underlying biology²². Whitfield *et al*²³ were amongst the first to implicate a large role for proliferation in the prognostic capacity of these gene expression-based classifiers. Subsequent research provides data supportive of this concept for breast cancer^{20,21,24,25}. Wirapati *et al*¹⁹ showed that discarding the proliferation-associated genes from a series of classifiers significantly deteriorated signature performance, whereas only focusing on this gene cluster dramatically improved prognosticity in several cases.

Proliferation signature

These data stimulated us to create an *in vitro*-based gene expression proliferation marker and test classification ability across different cancer types. High prognosticity of this proliferation signature was observed in five patient datasets representing breast, non-small-cell lung and renal cancer (chapter 2). In chapter 4 we subsequently show that conversion of our proliferation signature to a PCR-based test shows clinical potential in (early-stage) breast and non-small-cell lung cancer which can possibly be generalized across other cancer types. Markers that can be used across different tumor types might help elucidate what processes play a key role in tumorigenesis and tumor aggressiveness in general. However marker importance will probably not be equal for tumors of different origin site and different thresholds and cut-offs might apply.

Other proliferation markers

Proliferation substantially contributes to treatment response with both radiotherapy and chemotherapy²⁶. Subsequently, overall treatment time is an important contributor to outcome²⁷. Therefore, prognostic power of proliferation markers other than gene expression-based, like PCNA and Ki67 staining has extensively been studied^{28,29}. Although mitotic rate is included in the Nottingham grading system for breast cancer, in general published results of these single indicators vary widely across multiple cancer types: some studies show high prognostic power whereas others show none at all^{28,30-34}. More recently a set of meta-analyses on primary breast cancer reported consistently positive results for immunohistochemical evaluation of Ki67 as a prognostic marker³⁵⁻³⁷. Nevertheless, all these studies conclude that, at present, Ki67 cannot be used as a tool for risk group classification or therapy selection, in large part because of a lack of standard methodology and definition of threshold points. Studies have identified Ki67 to be involved in ribosomal RNA translation, however the exact functional role of Ki67 remains unclear^{38,39}. Ki67 is expressed at different levels throughout the cell cycle, with a peak expression during mitosis⁴⁰. Expression is generally lost in quiescent cells⁴¹. In general, scoring is based on the percentage of positive stained tumor cells. In the aforementioned meta-data analyses reported cut-off values ranged from ~3-40%. Until standard tissue preparation procedures, staining methods and cut-off points are defined, Ki67 cannot be introduced in a clinical set-up.

Since a measure of proliferation is included in the breast cancer grading system, and since evidence of the prognostic power of Ki67 is emerging, one could question the added value of new gene expression-based proliferation markers for breast cancer. Data from the two prognostic gene profiles^{6,9} currently approved by the FDA for early-stage breast cancer suggest otherwise. In multiple large patient cohorts, these classifiers outperform traditional clinical and histological parameters⁴²⁻⁴⁵. Prognostic capacity of both markers is significantly dependent on the contribution of proliferation-associated genes²⁴. In other cancer types far less data is available on the performance of proliferation markers, but results from gene expression microarray studies including our own show the potential of gene expression-based classifiers for proliferation in clinical practice^{46,47}. Evaluation of this PCR-based proliferation marker in large independent datasets, preferentially in a prospective setting, across different cancer types will elucidate the clinical applicability of this classifier.

Role of other processes in transcriptome-based prognosis

Although proliferation currently seems to be the driving force in prognostic power of patient data-based signatures it will not be the only process with prognostic and possible predictive capacity at the gene expression level. A comprehensive analysis of Reyal *et al*²⁵ encompassing several gene expression signatures and breast cancer datasets identified a set of modules enriched in the classification power of signatures which include genes involved in proliferation, immune response and RNA splicing. In addition various pathway-based signatures reflecting other cellular processes or micro-environmental features show potential. Multiple hypoxia or hypoxia-related gene sets have been created with prognostic power in different cancer types⁴⁸⁻⁵⁵. Further Carter *et al*⁷ reported a gene expression-based marker for chromosomal instability (CIN) that could stratify different patient datasets, comprising diverse cancer types, into groups with different survival properties.

Transcriptome-based markers for different cellular processes and micro-environmental features might thus elucidate independent prognostic power. This suggests that combining signatures could eventually lead to further stratification of patients with different survival properties and can aid in the selection of specific patient groups for therapies directed at these various processes. Previous studies have shown that combining the Wound signature⁵⁶ with either the 70-gene signature⁶, a hypoxia signature⁴⁸ or the invasiveness gene signature (IGS)¹¹ resulted

in the definition of three subclasses of patients with distinct differences in survival leading to improved prognostic performance compared to single marker performance^{11,57,58}.

Gene expression signatures in the clinic

A possible limitation to the application of gene expression-based classifiers is the intra-tumor heterogeneity of tumors. In general most of the gene expression microarray studies are performed on surgically treated patients and whole tumor tissues are available for profiling. In order to optimally apply these markers for patient treatment decisions it will be essential to know whether transcriptomic information on the whole tumor can be captured from biopsy material. This would allow information to be obtained at the diagnostic stage rather than during treatment, and thus to inform a larger fraction of patient care. In addition, impact on decision making in cancer treatment will be largest when non-surgically treated patients (i.e. those treated with radiotherapy alone) are also eligible for evaluation with these signatures.

The same issue manifested with the emergence of tissue microarray (TMA) technology for immunohistochemical assessment of biomarkers. Since then multiple studies were performed to assess the number of core-biopsies needed to obtain a representative view of a single tumor⁵⁹⁻⁶³. Now it is widely accepted that two to four samples from different tumor regions suffice to get a reliable estimation of the information in the whole specimen⁶⁴. However a recent study indicates that the number of tissue samples should be determined per individual biomarker for TMA evaluation⁶⁵. Similar studies should be performed to address this issue for gene expression data.

Validation in gene expression microarray research is not straightforward

Although clinical application of gene expression microarray technology is widely expected to introduce a new chapter in personalizing cancer treatment, progress has been slow. As indicated before skepticism regarding gene expression-based signatures has been growing since reports on failed re-analyses, disappointing validation results and lack of consistency started to appear^{13,15,16,66,67}. Both the high-dimensionality of the microarray data and the numerous methods available to process the data play an important role.

Multiple testing: the problem of high-dimensionality data

A large part of these negative results can be attributed to the fact that in microarray research we deal with high-dimensionality data coupled to insufficient statistical power. This problem is often referred to as ‘the curse of dimensionality’: the expression of (tens of) thousands of genes (independent variables) are simultaneously measured whereas the number of samples (dependent variables) is limited to maybe a few hundred. Without the use of proper statistical and computational methods to account for this difference in dimensionality there is a major risk of over-fitting, which can lead to the reporting of overoptimistic results^{13,14,68}.

When identifying genes differentially expressed between two patient groups traditional statistical methods such as the two-sample t-test will be challenged with serious multiplicity^{69,70}. For example, when taking a microarray with 25,000 genes and a standard measure for significance ($p < 0.05$) one will expect around 1,250 false discoveries. Various alternative methods have been suggested to better deal with this type of high-dimensionality data for gene selection⁷¹. Series of adjusted statistics, like Bonferroni adjustment⁷² and Benjamini and Hochberg correction⁷³, resampling techniques^{74,75} and dimensionality reduction methods^{76,77}, like principle component analysis (PCA)⁷⁸ and support vector machines (SVM)⁷⁹, have been introduced to control the level of false discoveries and elucidate the truly differentially expressed genes, which resulted in improved marker performance in validation data.

Influence of multiplicity in signature testing

Despite all statistically-based efforts to account for multiple testing, it is crucial to validate prognostic gene expression-based markers in independent data. Studied patient cohorts are merely small subsets of the total patient population of interest and are usually assembled at a single site. The diversity observed in the studied cohort is not representative of the variety in the total patient population for a number of reasons including differences in experimental procedure (RNA extraction, array batch effects) and epidemiological factors. On top of that, classification is often more relevant in a subgroup of the studied cohort, however subgroup analyses are largely underpowered owing to the small numbers of patients included. In addition the process of marker identification almost always includes a training step, true value of the created classifier can then only be

assessed by testing the marker in independent data. Prognostic power is determined with traditional survival statistics. Although only one or a few markers are typically validated, we show in chapter 5 that one has to deal with multiple testing problems at the signature testing phase. We propose a method to obtain an estimate of the performance of the signature in a specific dataset relative to an empirically-determined null distribution. We develop techniques to estimate this null distribution by evaluating randomly-generated gene signatures, and use these to confirm previous findings that the number of included samples dramatically affects the false discoveries showing the importance of large cohorts^{68,80,81}. Large differences in the performance of random gene sets were identified between datasets. Boutros *et al*⁸² recently used a similar permutation-based approach to evaluate their signature in multiple datasets.

In addition we show that once the number of validation datasets increases the multiple testing issue plays a smaller part (chapter 5). At the moment most published signatures are validated in a small number of datasets. Worse, for certain tumor types or specific treatments, for example patients treated with a new drug, the number of validation cohorts can be limited. Our new technique could strengthen the findings in these statistically-challenging but medically-critical cases, by adding an additional measure of classifier performance which accounts for the high data dimensionality.

Data Sensitivity to pre-processing

One of the other issues leading to inconsistent results and discrepancies in signature gene content is the large variety of microarray technologies, pre-processing methods and analysis techniques available in the field of gene expression microarray analyses⁷⁰. Initially, reports on inter-platform and cross-laboratory studies showed poor repeatability, reproducibility and comparability of microarray technology⁸³⁻⁸⁶. More recently, modern arrays and improved statistical methods have led to more concordant results: several studies show reasonable to good agreement of microarray data generated using different platforms or at different sites⁸⁷⁻⁸⁹. Nevertheless, when Ioannidis *et al*⁶⁶ tried to replicate data analyses from a series of published gene expression microarray studies, only a few analyses could be reproduced. This could mainly be attributed to the unavailability or incompleteness of available data and lack of reporting of analysis details. More transparent analyses descriptions and better data accessibility will thus lead to

increased data repeatability and will make evaluation of developed models by independent groups easier.

To systematically address issues regarding intra- and inter-platform reproducibility, the MicroArray Quality Control (MAQC) consortium, comprised of more than 50 organizations and 130 scientists, was established. In their first series of publications, consistent agreement was reported both for gene expression data from the same platform at different test sites and across data generated by different technologies⁹⁰⁻⁹⁵. In a large multi-site follow-up study, the group changed their focus from the technical aspects of gene expression microarray technology to the data analysis methods⁹⁶. Six datasets with different endpoints were analyzed by 36 independent teams to extract classification models. In total more than 30,000 models were generated using a wide variety of analytical methods. Overall models generated with different methods showed high agreement in performance. This confirms previous data from individual research groups demonstrating that the application of different algorithms yielded similar classifications^{12,17,82,97}. The studied endpoint and team competence largely influence model performance. They further emphasize the importance of applying the same processing pipeline to validation datasets as was used for the training datasets.

Studies evaluating different pre-processing schedules have shown that the use of different schemes greatly influence downstream analyses, which for example can result in completely different identified gene lists⁹⁸⁻¹⁰⁰. Coombes *et al*¹⁵ and Baggerly *et al*¹⁶ demonstrated that the use of different pre-processing resulted in failed reproduction of drug sensitivity predictions. These data could imply that diverging from the scheme used in an original report might lead to failed validation in subsequent studies and unnecessary negative conclusions on gene expression microarray technology. We systematically studied the influence of different pre-processing schedules on the performance of a previously published 3-gene lung cancer marker¹⁰¹ in a large independent microarray dataset⁴⁶ (chapter 7). When the prescribed pre-processing methodology was applied, the 3-gene marker successfully validated in this new cohort. However we show that minor alterations in pre-processing could change a successful biomarker into one indistinguishable from random chance. In a recent review, Subramanian and Simon⁶⁷ failed to validate two lung cancer signatures including the 3-gene marker in the same independent dataset and concluded that any of the NSCLC signatures available at the moment are ready for clinical application. We identified substantial differences in the

adopted pre-processing schedule used by Subramanian and Simon⁶⁷ compared to the one reported in the original study (chapter 6) and identify this as the cause of failed validation (chapter 7). The use of different pre-processing schedules resulted in changed classification status of a significant portion of patients (chapter 7). For now marker validation should follow a consistent pre-processing scheme. However, we hypothesize that this data pre-processing sensitivity can be used to improve biomarker robustness on a per-patient basis.

Standardizing the gene expression profiling pipeline for a specific marker, continued validation in new datasets and transparent descriptions of data analyses schedules will result in better gene expression-based markers and more rapid integration in clinical settings.

Future perspectives

Accelerating clinical utility of high-throughput gene expression profiling

The first mRNA-based classifiers are currently being tested in large clinical prospective trials¹⁰²⁻¹⁰⁴. As a result, the use of microarrays for high-throughput gene-expression profiling shows significant potential for predicting both prognosis and response to therapy. Since the introduction of microarray-based technology, many efforts have been made to revise study design, pre-processing and analysis methodology to improve clinical utility¹⁰⁵. Continued advances in computational and experimental approaches will make it increasingly practical to translate microarray studies to patient settings.

One important issue that remains to be addressed is that most markers are evaluated in a relative fashion. The earliest gene expression microarray studies used a subset of genes to cluster patients in different groups^{5,106,107}. This approach has several disadvantages, for example unsupervised clustering could be unstable in the presence of outliers, and the employment of a different algorithm might result in completely different patient groups, making this a subjective analysis strategy¹⁰⁸. Recent studies have therefore moved towards more quantitative evaluations, like assigning scores derived from expression of each genes in the classifier. However in most cases these scores are used to arbitrarily group patients, for example by dichotomizing the median score within a population. Converting these markers to

absolute measures to provide rigorous estimates of patient risk, and the establishment of the associated classification thresholds and cut-off values will be essential for introduction into patient decision making.

Another important requirement for the success of future studies is complete data and analysis procedure reporting^{66,67,105}. Transparent reporting of computational approaches makes it easier for external parties to evaluate published gene signatures in their own datasets. In addition availability of more raw datasets will increase the amount of data that can be used for validation purposes. These will lead to an increase in the rate at which a classifier will be accepted, and may indeed increasingly become a requirement for approval by governments and insurance companies.

Next-generation sequencing (NGS) technology

Gene expression profiling accounts for the majority of publications on markers identified with high throughput technologies. However, recent advances in next-generation sequencing (NGS) have enormously decreased the costs of high-throughput sequencing, making this technique practical for large cohorts for the first time¹⁰⁹. Transcriptome or exome NGS will eventually overtake application of gene expression microarrays for several reasons. First of all NGS provides exquisite sensitivity and resolution far beyond that of other high-throughput technologies^{110,111}. Second, exome and transcriptome NGS provides much more information than microarrays¹¹². Data can also be used to identify gene fusions¹¹³, alternative splicing variants¹¹⁴ and possibly different types of mutations¹¹⁵. Third, analyses are not limited to the transcripts available on a gene chip, but can also detect novel transcripts^{116,117}. Thus with the emergency of affordable NGS, enormous datasets will become available and the analyses of these data will become a key and challenging task^{109,118}. Tools to analyze this type of data have started to emerge but, are currently very limited in scope and validation¹¹⁹.

The increased accessibility of NGS data further provides the possibility to study much more than only the transcriptome¹²⁰. Whole genome sequencing will generate a complete spectrum of the mutations, amplifications and silencing present in a tumor sample¹¹². For this type of NGS it will be most important to identify what genetic alterations are driving tumorigenesis and tumor progression and which ones are secondary or passenger variations¹⁰⁹. Targeted NGS, like chromosome immune-precipitation (ChIP) followed by high-throughput sequencing

procures data on the epigenome, like DNA methylation or histone acetylation, and DNA-protein interactions¹²¹. Combining different NGS types will in the end dramatically expand our understanding of cancer as a disease and ultimately result in better diagnostics, new and improved targeted treatments and increased survival.

The need for multi-factorial decision tools

Due to the complexity of tumor biology it is likely that only the employment of multi-factorial markers will lead to truly individualized therapy decisions. To understand complex cancer biology it will be key to elucidate the molecular events at different levels like genomics, transcriptomics and proteomics and to combine these into a unified model of the effects and implications on functional molecular pathway and treatment outcome. Ultimately information from multiple modalities like patient characteristics, clinical and staging data, molecular markers, imaging data and treatment information will be integrated in decision support systems (DSSs) to optimize treatment for each individual patient.

We are still far from realizing such extensive DSSs to integrate high-throughput data with clinical and imaging parameters. Most studies so far focus on the use of a single modality to define novel molecular prognosticators; follow-up studies are needed to address whether these markers in combination with other parameters really increase clinical utility.

References

1. Borjesson, P.K., *et al.* Performance of immuno-positron emission tomography with zirconium-89-labeled chimeric monoclonal antibody U36 in the detection of lymph node metastases in head and neck cancer patients. *Clin Cancer Res* **12**, 2133-2140 (2006).
2. van Baardwijk, A., *et al.* The maximum uptake of (18)F-deoxyglucose on positron emission tomography scan correlates with survival, hypoxia inducible factor-1alpha and GLUT-1 in non-small cell lung cancer. *Eur J Cancer* **43**, 1392-1398 (2007).
3. Mack, P.C., *et al.* Lower osteopontin plasma levels are associated with superior outcomes in advanced non-small-cell lung cancer patients receiving platinum-based chemotherapy: SWOG Study S0003. *J Clin Oncol* **26**, 4771-4776 (2008).
4. Cocco, E., *et al.* Serum amyloid A: a novel biomarker for endometrial cancer. *Cancer* **116**, 843-851 (2010).
5. Perou, C.M., *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747-752 (2000).
6. van 't Veer, L.J., *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530-536 (2002).
7. Carter, S.L., Eklund, A.C., Kohane, I.S., Harris, L.N. & Szallasi, Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* **38**, 1043-1048 (2006).
8. Chen, H.Y., *et al.* A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* **356**, 11-20 (2007).
9. Paik, S., *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* **351**, 2817-2826 (2004).
10. Chung, C.H., *et al.* Increased epidermal growth factor receptor gene copy number is associated with poor prognosis in head and neck squamous cell carcinomas. *J Clin Oncol* **24**, 4170-4176 (2006).
11. Liu, R., *et al.* The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* **356**, 217-226 (2007).
12. Ein-Dor, L., Zuk, O. & Domany, E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci U S A* **103**, 5923-5928 (2006).
13. Michiels, S., Koscielny, S. & Hill, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* **365**, 488-492 (2005).
14. Dupuy, A. & Simon, R.M. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* **99**, 147-157 (2007).
15. Coombes, K.R., Wang, J. & Baggerly, K.A. Microarrays: retracing steps. *Nat Med* **13**, 1276-1277; author reply 1277-1278 (2007).
16. Baggerly, K.A., Coombes, K.R. & Neeley, E.S. Run batch effects potentially compromise the usefulness of genomic signatures for ovarian cancer. *J Clin Oncol* **26**, 1186-1187; author reply 1187-1188 (2008).
17. Fan, C., *et al.* Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* **355**, 560-569 (2006).
18. Thomassen, M., *et al.* Comparison of gene sets for expression profiling: prediction of metastasis from low-malignant breast cancer. *Clin Cancer Res* **13**, 5355-5360 (2007).
19. Wirapati, P., *et al.* Meta-analysis of gene expression profiles in breast cancer: toward a unified

- understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* **10**, R65 (2008).
20. Desmedt, C., *et al.* Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res* **14**, 5158-5165 (2008).
 21. Haibe-Kains, B., Desmedt, C., Sotiriou, C. & Bontempi, G. A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? *Bioinformatics* **24**, 2200-2208 (2008).
 22. Shi, W., *et al.* Functional analysis of multiple genomic signatures demonstrates that classification algorithms choose phenotype-related genes. *Pharmacogenomics J* **10**, 310-323 (2010).
 23. Whitfield, M.L., George, L.K., Grant, G.D. & Perou, C.M. Common markers of proliferation. *Nat Rev Cancer* **6**, 99-106 (2006).
 24. Weigelt, B., Baehner, F.L. & Reis-Filho, J.S. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. *J Pathol* **220**, 263-280 (2010).
 25. Reyat, F., *et al.* A comprehensive analysis of prognostic signatures reveals the high predictive capacity of the proliferation, immune response and RNA splicing modules in breast cancer. *Breast Cancer Res* **10**, R93 (2008).
 26. Bourhis, J., *et al.* Hyperfractionated or accelerated radiotherapy in head and neck cancer: a meta-analysis. *Lancet* **368**, 843-854 (2006).
 27. De Ruysscher, D., *et al.* Time between the first day of chemotherapy and the last day of chest radiation is the most important predictor of survival in limited-disease small-cell lung cancer. *J Clin Oncol* **24**, 1057-1063 (2006).
 28. Pich, A., Chiusa, L. & Navone, R. Prognostic relevance of cell proliferation in head and neck tumors. *Ann Oncol* **15**, 1319-1329 (2004).
 29. Stuart-Harris, R., Caldas, C., Pinder, S.E. & Pharoah, P. Proliferation markers and survival in early breast cancer: a systematic review and meta-analysis of 85 studies in 32,825 patients. *Breast* **17**, 323-334 (2008).
 30. Begg, A.C., *et al.* The value of pretreatment cell kinetic parameters as predictors for radiotherapy outcome in head and neck cancer: a multicenter analysis. *Radiother Oncol* **50**, 13-23 (1999).
 31. Daniels, J.M., *et al.* Mitotic index does not predict prognosis in stage IA non-small cell lung cancer. *Lung Cancer* **38**, 163-167 (2002).
 32. Caly, M., *et al.* Analysis of correlation between mitotic index, MIB1 score and S-phase fraction as proliferation markers in invasive breast carcinoma. Methodological aspects and prognostic value in a series of 257 cases. *Anticancer Res* **24**, 3283-3288 (2004).
 33. Shomori, K., *et al.* Geminin, Ki67, and minichromosome maintenance 2 in gastric hyperplastic polyps, adenomas, and intestinal-type carcinomas: pathobiological significance. *Gastric Cancer* **13**, 177-185 (2010).
 34. Noordhuis, M.G., *et al.* Prognostic cell biological markers in cervical cancer patients primarily treated with (chemo)radiation: a systematic review. *Int J Radiat Oncol Biol Phys* **79**, 325-334 (2011).
 35. Colozza, M., *et al.* Proliferative markers as prognostic and predictive tools in early breast cancer: where are we now? *Ann Oncol* **16**, 1723-1739 (2005).
 36. de Azambuja, E., *et al.* Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. *Br J Cancer* **96**, 1504-1513 (2007).

37. Yerushalmi, R., Woods, R., Ravdin, P.M., Hayes, M.M. & Gelmon, K.A. Ki67 in breast cancer: prognostic and predictive potential. *Lancet Oncol* **11**, 174-183 (2010).
38. Bullwinkel, J., *et al.* Ki-67 protein is associated with ribosomal RNA transcription in quiescent and proliferating cells. *J Cell Physiol* **206**, 624-635 (2006).
39. Rahmanzadeh, R., Huttmann, G., Gerdes, J. & Scholzen, T. Chromophore-assisted light inactivation of pKi-67 leads to inhibition of ribosomal RNA synthesis. *Cell Prolif* **40**, 422-430 (2007).
40. Gerdes, J., Schwab, U., Lemke, H. & Stein, H. Production of a mouse monoclonal antibody reactive with a human nuclear antigen associated with cell proliferation. *Int J Cancer* **31**, 13-20 (1983).
41. Beresford, M.J., Wilson, G.D. & Makris, A. Measuring proliferation in breast cancer: practicalities and applications. *Breast Cancer Res* **8**, 216 (2006).
42. Buyse, M., *et al.* Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* **98**, 1183-1192 (2006).
43. Mook, S., *et al.* The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1-3 positive lymph nodes in an independent validation study. *Breast Cancer Res Treat* **116**, 295-302 (2009).
44. Goldstein, L.J., *et al.* Prognostic utility of the 21-gene assay in hormone receptor-positive operable breast cancer compared with classical clinicopathologic features. *J Clin Oncol* **26**, 4063-4071 (2008).
45. Habel, L.A., *et al.* A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res* **8**, R25 (2006).
46. Shedden, K., *et al.* Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* **14**, 822-827 (2008).
47. Cuzick, J., *et al.* Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *Lancet Oncol* **12**, 245-255 (2011).
48. Chi, J.T., *et al.* Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. *PLoS Med* **3**, e47 (2006).
49. Sung, F.L., *et al.* Genome-wide expression analysis using microarray identified complex signaling pathways modulated by hypoxia in nasopharyngeal carcinoma. *Cancer Lett* **253**, 74-88 (2007).
50. Winter, S.C., *et al.* Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. *Cancer Res* **67**, 3441-3449 (2007).
51. Hu, Z., *et al.* A compact VEGF signature associated with distant metastases and poor outcomes. *BMC Med* **7**, 9 (2009).
52. Fardin, P., *et al.* A biology-driven approach identifies the hypoxia gene signature as a predictor of the outcome of neuroblastoma patients. *Mol Cancer* **9**, 185 (2010).
53. van Malenstein, H., *et al.* A seven-gene set associated with chronic hypoxia of prognostic importance in hepatocellular carcinoma. *Clin Cancer Res* **16**, 4278-4288 (2010).
54. Buffa, F.M., Harris, A.L., West, C.M. & Miller, C.J. Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *Br J Cancer* **102**, 428-435 (2010).
55. Seigneuric, R., *et al.* Impact of supervised gene signatures of early hypoxia on patient survival. *Radiother Oncol* **83**, 374-382 (2007).
56. Chang, H.Y., *et al.* Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* **2**, E7 (2004).
57. Chang, H.Y., *et al.* Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* **102**, 3738-3743 (2005).

58. Nuyten, D.S., Hastie, T., Chi, J.T., Chang, H.Y. & van de Vijver, M.J. Combining biological gene expression signatures in predicting outcome in breast cancer: An alternative to supervised classification. *Eur J Cancer* **44**, 2319-2329 (2008).
59. Camp, R.L., Charette, L.A. & Rimm, D.L. Validation of tissue microarray technology in breast carcinoma. *Lab Invest* **80**, 1943-1949 (2000).
60. Hoos, A., *et al.* Validation of tissue microarrays for immunohistochemical profiling of cancer specimens using the example of human fibroblastic tumors. *Am J Pathol* **158**, 1245-1251 (2001).
61. Nocito, A., *et al.* Microarrays of bladder cancer tissue are highly representative of proliferation index and histological grade. *J Pathol* **194**, 349-357 (2001).
62. Rubin, M.A., Dunn, R., Strawderman, M. & Pienta, K.J. Tissue microarray sampling strategy for prostate cancer biomarker analysis. *Am J Surg Pathol* **26**, 312-319 (2002).
63. Gulmann, C., Butler, D., Kay, E., Grace, A. & Leader, M. Biopsy of a biopsy: validation of immunoprofiling in gastric cancer biopsy tissue microarrays. *Histopathology* **42**, 70-76 (2003).
64. Chen, W. & Foran, D.J. Advances in cancer tissue microarray technology: Towards improved understanding and diagnostics. *Anal Chim Acta* **564**, 74-81 (2006).
65. Eckel-Passow, J.E., *et al.* Tissue microarrays: one size does not fit all. *Diagn Pathol* **5**, 48 (2010).
66. Ioannidis, J.P., *et al.* Repeatability of published microarray gene expression analyses. *Nat Genet* **41**, 149-155 (2009).
67. Subramanian, J. & Simon, R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Natl Cancer Inst* **102**, 464-474 (2010).
68. Ntzani, E.E. & Ioannidis, J.P. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* **362**, 1439-1444 (2003).
69. Miller, L.D., *et al.* Optimal gene expression analysis by microarrays. *Cancer Cell* **2**, 353-361 (2002).
70. Jeanmougin, M., *et al.* Should we abandon the t-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies. *PLoS One* **5**, e12336 (2010).
71. Farcomeni, A. A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. *Stat Methods Med Res* **17**, 347-388 (2008).
72. Dunn, O.J. Multiple Comparisons Among Means. *Journal of the American Statistical Association* **56**, 52-64 (1961).
73. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **57**, 289-300 (1995).
74. Jensen, S.T., Soi, S. & Wang, L.S. A Bayesian approach to efficient differential allocation for resampling-based significance testing. *BMC Bioinformatics* **10**, 198 (2009).
75. Zhang, W., Robbins, K., Wang, Y., Bertrand, K. & Rekaya, R. A jackknife-like method for classification and uncertainty assessment of multi-category tumor samples using gene expression information. *BMC Genomics* **11**, 273 (2010).
76. Bartenhagen, C., Klein, H.U., Ruckert, C., Jiang, X. & Dugas, M. Comparative study of unsupervised dimension reduction techniques for the visualization of microarray gene expression data. *BMC Bioinformatics* **11**, 567 (2010).
77. Huang, J., Fang, H. & Fan, X. Decision forest for classification of gene expression data. *Comput Biol Med* **40**, 698-704 (2010).
78. Ma, S. & Dai, Y. Principal component analysis based methods in bioinformatics studies. *Brief Bioinform* (2011).

79. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene selection for cancer classification using Support Vector Machines. *Machine Learning* **46**, 389-422 (2002).
80. Zien, A., Fluck, J., Zimmer, R. & Lengauer, T. Microarrays: how many do you need? *J Comput Biol* **10**, 653-667 (2003).
81. Popovici, V., et al. Effect of training-sample size and classification difficulty on the accuracy of genomic predictors. *Breast Cancer Res* **12**, R5 (2010).
82. Boutros, P.C., et al. Prognostic gene signatures for non-small-cell lung cancer. *Proc Natl Acad Sci U S A* **106**, 2824-2828 (2009).
83. Kuo, W.P., Jenssen, T.K., Butte, A.J., Ohno-Machado, L. & Kohane, I.S. Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* **18**, 405-412 (2002).
84. Tan, P.K., et al. Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* **31**, 5676-5684 (2003).
85. Miklos, G.L. & Maleszka, R. Microarray reality checks in the context of a complex disease. *Nat Biotechnol* **22**, 615-621 (2004).
86. Frantz, S. An array of problems. *Nat Rev Drug Discov* **4**, 362-363 (2005).
87. Petersen, D., et al. Three microarray platforms: an analysis of their concordance in profiling gene expression. *BMC Genomics* **6**, 63 (2005).
88. Dobbin, K.K., et al. Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. *Clin Cancer Res* **11**, 565-572 (2005).
89. Larkin, J.E., Frank, B.C., Gavras, H., Sultana, R. & Quackenbush, J. Independence and reproducibility across microarray platforms. *Nat Methods* **2**, 337-344 (2005).
90. Shi, L., et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* **24**, 1151-1161 (2006).
91. Patterson, T.A., et al. Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nat Biotechnol* **24**, 1140-1150 (2006).
92. Guo, L., et al. Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nat Biotechnol* **24**, 1162-1169 (2006).
93. Canales, R.D., et al. Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotechnol* **24**, 1115-1122 (2006).
94. Shippy, R., et al. Using RNA sample titrations to assess microarray platform performance and normalization techniques. *Nat Biotechnol* **24**, 1123-1131 (2006).
95. Tong, W., et al. Evaluation of external RNA controls for the assessment of microarray performance. *Nat Biotechnol* **24**, 1132-1139 (2006).
96. Shi, L., et al. The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat Biotechnol* **28**, 827-838 (2010).
97. Ein-Dor, L., Kela, I., Getz, G., Givol, D. & Domany, E. Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics* **21**, 171-178 (2005).
98. Irizarry, R.A., et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-264 (2003).
99. Lim, W.K., Wang, K., Lefebvre, C. & Califano, A. Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks. *Bioinformatics* **23**, i282-288 (2007).
100. Hoffmann, R., Seidl, T. & Dugas, M. Profound effect of normalization on detection of differentially

- expressed genes in oligonucleotide microarray data analysis. *Genome Biol* **3**, RESEARCH0033 (2002).
101. Lau, S.K., *et al.* Three-gene prognostic classifier for early-stage non small-cell lung cancer. *J Clin Oncol* **25**, 5562-5569 (2007).
 102. Sparano, J.A. TAILORx: trial assigning individualized options for treatment (Rx). *Clin Breast Cancer* **7**, 347-350 (2006).
 103. Bogaerts, J., *et al.* Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. *Nat Clin Pract Oncol* **3**, 540-551 (2006).
 104. Cardoso, F., *et al.* Clinical application of the 70-gene profile: the MINDACT trial. *J Clin Oncol* **26**, 729-735 (2008).
 105. Tillinghast, G.W. Microarrays in the clinic. *Nat Biotechnol* **28**, 810-812 (2010).
 106. Bittner, M., *et al.* Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* **406**, 536-540 (2000).
 107. Alizadeh, A.A., *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503-511 (2000).
 108. Simon, R., Radmacher, M.D., Dobbin, K. & McShane, L.M. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* **95**, 14-18 (2003).
 109. Meyerson, M., Gabriel, S. & Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* **11**, 685-696 (2010).
 110. Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* **18**, 1509-1517 (2008).
 111. Morrissy, A.S., *et al.* Next-generation tag sequencing for cancer gene expression profiling. *Genome Res* **19**, 1825-1835 (2009).
 112. Mardis, E.R. & Wilson, R.K. Cancer genome sequencing: a review. *Hum Mol Genet* **18**, R163-168 (2009).
 113. Maher, C.A., *et al.* Transcriptome sequencing to detect gene fusions in cancer. *Nature* **458**, 97-101 (2009).
 114. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).
 115. Shah, S.P., *et al.* Mutation of FOXL2 in granulosa-cell tumors of the ovary. *N Engl J Med* **360**, 2719-2729 (2009).
 116. Sultan, M., *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956-960 (2008).
 117. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628 (2008).
 118. Koboldt, D.C., Ding, L., Mardis, E.R. & Wilson, R.K. Challenges of sequencing human genomes. *Brief Bioinform* **11**, 484-498 (2010).
 119. Ding, L., Wendl, M.C., Koboldt, D.C. & Mardis, E.R. Analysis of next-generation genomic data in cancer: accomplishments and challenges. *Hum Mol Genet* **19**, R188-196 (2010).
 120. Metzker, M.L. Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46 (2010).
 121. Hoffman, B.G. & Jones, S.J. Genome-wide identification of DNA-protein interactions using chromatin immunoprecipitation coupled with flow cell sequencing. *J Endocrinol* **201**, 1-13 (2009).

Chapter 9

Summary

Summary

One of the major focuses in cancer research is the identification of novel biomarkers and therapeutic targets to improve treatment outcomes. There is a significant need to enhance both the accuracy and precision of risk prediction and therapy selection. Advances in both molecular and clinical cancer research have improved prognosis and reduced side-effects. Nonetheless, available therapy remains ineffective in a large fraction of patients. New biomarkers could help predict which patients would benefit from a treatment.

With the advent of high-throughput technologies it became possible to study enormous numbers of parameters simultaneously, opening up a new era of possibilities to identify novel markers much faster than previously possible.

One of these novel technologies, gene-expression microarrays have been widely used in cancer research for the last decade. Measuring the expression of thousands of genes in parallel has the potential to answer many questions that were previously impossible to resolve. This type of transcriptomic profiling has, amongst others, led to the identification of dysregulated pathways in cancer. Furthermore, multi-gene markers (so-called gene signatures) have been created that show the promise to improve risk prediction and therapy selection. Despite progress in this area, introduction into clinical practice has been slow. Only two breast cancer gene signatures are currently tested in clinical trials. Reports on minimal gene-wise overlap, disappointing validation in external data and lack of consistency have negatively impacted the field.

The first part of this thesis focuses on the creation of a robust, biology-based gene signature for multiple cancer types. In the second part of this thesis we elaborate on the data analysis of microarray experiments.

Proliferation signature

Recently it has been reported that, although gene-wise overlap is poor, classification results from different signatures show high agreement. It is thought that this phenomenon reflects common underlying biology represented by each classifier in a different way. In breast cancer, more and more data suggests that cellular proliferation plays a major role in the prognosticity of gene signatures.

To clarify the significance of proliferation in prognostic breast cancer signatures we created a gene expression-based marker reflecting cellular proliferation (**chapter 2**).

Two *in vitro* gene expression datasets were combined to extract this signature: genes were selected that showed a cycling expression pattern after synchronization in one dataset and responded to serum stimulation in the other. This 104-gene proliferation signature had high prognostic power in different breast cancer datasets. We further hypothesized that this marker could also be used for prognostic purposes in other cancer sites and validated it across multiple cancer types. A further independent validation of our proliferation signature was performed by relating gene expression to other proliferation measures in both patient and xenograft material (**chapter 3**).

Since the clinical utility of microarrays can be controversial, we sought to evaluate our classifier using an independent technique. In **chapter 4** the number of genes in the signature was computationally-reduced in an information-preserving manner, followed by experimental evaluation using RT-PCR. This reduced marker was successfully tested in two large patient microarray meta-datasets. Subsequently, this refined marker was validated with PCR *in vitro*, *in vivo* and in an independent primary breast cancer cohort.

Challenges in microarray data analysis

Several aspects of high-throughput gene expression profiling involve major technical challenges. The first lies in the dimensionality of a microarray experiment, where thousands of genes are simultaneously measured in a relatively small number of samples. Without the use of proper statistical and computational methods to account for this “curse of dimensionality” there is a major risk of over-fitting, which can lead to the reporting of over-optimistic and non-generalizable results. Increasingly sophisticated methods are being developed to handle this problem when identifying interesting genes and signatures from microarray data. This has clearly led to an improvement in quality of the reported classifiers. Less notice of the multiple testing issue is taken once a signature is created. Nevertheless, despite all efforts to account for multiple testing and over-fitting, it is crucial to independently validate prognostic gene expression-based. At this stage standard survival statistics are generally used to address prognosticity.

In **chapter 5** we clearly shown that also in this phase great care must be taken. By testing sets of randomly-generated signatures in multiple independent datasets we demonstrate that many gene sets reach statistical significance. In some datasets the proportion of significant gene-sets reached dramatic levels, indicating that a single,

dataset-invariant threshold for significance is inappropriate. Based on this permutation study, we suggest a method to test signature performance to control for spurious random findings. In addition, our data demonstrates that evaluating a signature in multiple independent datasets attenuates the multiple testing issue.

Another reason for discrepancies and inconsistencies in reported studies is the enormous diversity of available microarray platforms and pre-processing techniques. Variations in either of these parameters can result in identification of different genes and signatures. Many previously reported failed data replications can be attributed to unavailability or incompleteness of data or analysis details.

We therefore believe it is essential to follow the original data handling pipeline as precisely as possible (**chapter 6**), as failure to do so may in part explain why independent validation of signatures by other research groups often fails. Although in most of the cases the statistical methods are properly followed, far less attention is usually paid to data pre-processing. We sought to clarify the importance of standardizing the pre-processing schedule for gene expression-based classifiers. Two existing non-small-cell lung cancer signatures were tested in an independent dataset pre-processed with 24 different methods. The data in **chapter 7** clearly demonstrates that even small changes in pre-processing could change a successful marker into one indistinguishable from chance.

Our results do not only reveal the sensitivity of marker performance to the applied pre-processing, but also lead to the report of a new validation of two non-small-cell lung cancer markers in a new, large patient cohort. Interestingly, this study revealed that when focusing on those patients for whom different pre-processing schemes agree prognostic capability increased. This feature might be exploited to improve classifier robustness.

In conclusion, we created a signature for cellular proliferation, which demonstrated high prognostic power in multiple cancer types. We successfully translated this marker to a PCR-based classifier, which will increase clinical utility. In the second part of the thesis we identify several key issues that when kept in mind will improve reported classifier validity. Two novel methods for assessing signature robustness were developed. Taken together, these results make concrete and practical steps towards speeding the integration of molecular diagnostics with clinical practice.

Nederlandse samenvatting

Samenvatting

Het identificeren van nieuwe biomarkers en het stellen van therapeutische doelen ter verbetering van de behandeling is een van de speerpunten van kankeronderzoek. Het is duidelijk dat er nauwkeurigere methoden voor het voorspellen van het behandelingsresultaat en het bepalen van de optimale behandeling per individu nodig zijn. Vooruitgang op het gebied van onderzoek en verbetering van de bestaande behandelmethodes hebben zeer zeker een positief effect gehad op de prognose van patiënten. Echter, voor een groot deel van de patiënten blijven beschikbare therapieën inefficiënt of zelfs onbeduidend.

Met de opkomst van 'high-throughput' technieken is het mogelijk geworden om een zeer groot aantal parameters tegelijkertijd te bestuderen. Door deze ontwikkeling kunnen nieuwe markers vele malen sneller worden geïdentificeerd dan voorheen mogelijk was.

Een van deze nieuwe methoden, genaamd gen-expressie microarray technologie, wordt reeds veelvuldig gebruikt binnen het kankeronderzoek. Met deze techniek kan de expressie van duizenden genen tegelijkertijd worden gemeten, waardoor het mogelijk is om vele vraagstukken te ontrafelen die voorheen onmogelijk te beantwoorden waren. Er zijn sinds de introductie van deze methode verscheidene biologische processen gevonden die ontregeld zijn binnen kankercellen. Daarnaast worden gen-expressie microarrays veelvuldig gebruikt om multi-gen markers te genereren. Een zodanig samengestelde set genen heeft de potentie om te voorspellen welke patiënten goed op een behandeling zullen reageren en welke een slechtere prognose hebben. Het idee is dat deze markers uiteindelijk zullen bijdragen aan het bepalen van de meest optimale behandeling voor een patiënt.

Ondanks de voortgang in onderzoek op dit gebied verloopt de introductie van gen-expressie microarrays en geïdentificeerde multi-gen markers in een klinische setting traag. Momenteel zijn er slechts twee grote klinische studies gestart waarin multi-gen markers voor borstkanker worden getest. Rapportage van minimale gen overlap tussen markers, teleurstellende validatie resultaten in externe data en een gebrek aan consistentie dragen hiertoe bij.

Het eerste gedeelte van deze thesis behandelt de ontwikkeling van een robuuste multi-gen marker die mogelijk voor meerdere kankersoorten kan worden toegepast.

In het tweede deel wordt dieper ingegaan op de analyse van gen-expressie microarray data en wat de invloed hiervan is op marker resultaten.

Multi-gen marker voor proliferatie

Uit recent onderzoek blijkt ondanks dat er een gebrek aan overlap tussen multi-gen markers is, dat de patiënt classificaties met deze markers zeer goed overeenstemmen. Er wordt gedacht dat ongeacht of deze multi-gen markers qua inhoud verschillen of niet, ze gelijkwaardige biologie representeren. Verder wijst steeds meer onderzoek naar borstkanker erop dat proliferatie een belangrijke rol speelt ten aanzien van de prognostische waarde van multi-gen markers.

Om deze redenen hebben we in **hoofdstuk 2** een multi-gen marker ontwikkeld voor proliferatie. Onze verwachting is dat deze marker niet alleen van prognostische waarde zal zijn voor borstkanker patiënten, maar ook toepasbaar zal zijn voor andere kankersoorten. Om deze set van genen te creëren is informatie uit twee *in vitro* microarray datasets gecombineerd. In totaal werden 104 genen geselecteerd die een cyclisch expressie patroon volgden na synchronisatie in de ene dataset en een reactie vertoonden op stimulatie met serum in een andere dataset. Deze multi-gen proliferatie marker heeft een hoge prognostische waarde in verschillende patiënten microarray datasets. Een tweede validatie werd uitgevoerd door expressie van de proliferatie-geassocieerde genen te correleren aan andere metingen van proliferatie materiaal van patiënten en xenografts (**hoofdstuk 3**).

Aangezien de introductie van microarrays in een klinische setting nog moeilijk verloopt, hebben we gezocht naar een alternatieve methode om de proliferatie marker te bekijken. In **hoofdstuk 4** hebben we het aantal genen in de multi-gen marker verkleind, zodat het mogelijk werd om deze te evalueren met een onafhankelijke techniek: PCR. De prognostische waarde van deze nieuwe marker werd succesvol getest in twee grote microarray meta-datasets van patiënten. Daaropvolgend werd de aangepaste marker *in vitro*, *in vivo* en in een onafhankelijke patiënten dataset gevalideerd met PCR.

Microarray data analyse

Verschillende aspecten van 'high-throughput' methodes gaan gepaard met aanzienlijke technische vereisten. De dimensionaliteit van microarray data, waarbij expressie van duizenden genen tegelijkertijd wordt gemeten in een relatief klein aantal monsters, brengt een aantal uitdagingen met zich mee wat betreft de data-

analyse. Zonder het gebruik van geschikte statistische methoden en computer algoritmes bestaat er een groot risico van over-fitting, dat kan leiden tot rapportage van over-optimistische resultaten. Er worden steeds meer geavanceerde procedures ontwikkeld om hiervoor te corrigeren tijdens het identificeren van interessante genen en het afleiden van multi-gen markers. Dit heeft zeker effect gehad op de kwaliteit van gegenereerde markers. Er wordt minder aandacht besteed aan dit probleem nadat een nieuwe marker is gecreërd, het is echter essentieel om prognostische gen-expressie gebaseerde markers te valideren op onafhankelijke data. Om de prognostische waarde van deze markers te bepalen worden meestal standaard methodes voor overlevings statistiek toegepast.

In **hoofdstuk 5** tonen we aan dat ook in dit stadium rekening gehouden moet worden met de dimensionaliteit van de data. Door het genereren en testen van random multi-gen markers, werd duidelijk dat een groot aantal van deze willekeurige markers statistisch significante resultaten opleverden. Afhankelijk van de dataset bereikte dit aantal een dramatisch hoog niveau, waardoor het gebruik van een enkele drempel voor significantie niet geschikt is voor dit soort analyses. Om te controleren voor random bevindingen stellen we een methode voor om multi-gen markers te evalueren gebaseerd op deze permutatie studie. Daarnaast wijst de data in deze studie erop dat bij het testen van een marker in meerdere onafhankelijke datasets het van minder groot belang is om hier rekening mee te houden.

Een andere reden voor gerapporteerde tegenstrijdigheden en instabiliteit van multi-gen markers is het grote scala aan beschikbare microarray types, data voorbewerkingstechnieken en analyse methoden. Veranderingen in een van deze parameters resulteert in de identificatie van andere genen en dus ook in uiteenlopende multi-gen markers. Eerdere beschrijvingen over mislukte data replicatie konden meestal worden toegeschreven aan onvolledige of niet beschikbare ruwe data of analyse details.

Voor externe validatie denken we dat het essentieel is om de data bewerking zo precies mogelijk te volgen als omschreven in de oorspronkelijke studie (**hoofdstuk 6**). Onafhankelijke validatie van multi-gen markers geïdentificeerd door andere onderzoeksgroepen leveren doorgaans negatieve resultaten op. Hoewel in de meeste gevallen de analyse methoden tot in detail worden gevolgd, wordt er veel minder aandacht besteedt aan de data voorbewerking. Om het belang van

standaardisering van data voorbereiding te verduidelijken, werden twee bestaande multi-gen markers voor niet-kleincellig long carcinoma getest in een onafhankelijke dataset die met 24 verschillende schema's werd voorbereid. Uit de data in **hoofdstuk 7** blijkt dat zelfs kleine aanpassingen in een voorbereidings schema een succesvolle marker kunnen veranderen in een marker die niet te onderscheiden is van toeval.

In dit hoofdstuk laten we zien dat de prognostische waarde van een marker zeer gevoelig is voor de gebruikte voorbereidings techniek. Ook valideren we twee multi-gen markers voor niet-kleincellig long carcinoma in een nieuwe, grote dataset. Een andere interessante bevinding van deze studie is dat wanneer alleen patiënten in acht worden genomen waarvoor de verscheidene voorbereidingen een gelijke classificatie geven, de prognostische waarde sterk toeneemt. Deze eigenschap zou in de toekomst gebruikt kunnen worden om marker robuustheid te verbeteren.

Concluderend, omschrijven we in deze thesis de ontwikkeling van een multi-gen marker voor proliferatie, die een hoge prognostische waarde heeft in meerdere kankertypes. Deze marker werd succesvol omgezet in een PCR-gebaseerde marker, waardoor de toepasbaarheid in een klinische setting vergemakkelijkt zal worden. In het tweede deel van deze thesis identificeren we een aantal belangrijke punten waar rekening mee gehouden dient te worden bij de analyse van microarray data. Twee nieuwe methoden werden ontwikkeld om de robuustheid van een marker te bepalen. Dit zal bijdragen aan het verbeteren van de kwaliteit van multi-gen markers en leiden tot een snellere vertaling naar de kliniek.

Acknowledgements / Dankwoord

Acknowledgements / Dankwoord

Het schrijven van dit dankwoord is de laatste stap bij de afronding van dit proefschrift. Ik kijk met veel plezier terug op de afgelopen jaren van onderzoek die hier voor jullie liggen. Graag wil ik iedereen bedanken die -in welke vorm dan ook- heeft bijgedragen aan dit boekje. Ik had dit niet voor elkaar kunnen krijgen zonder de hulp en aanmoediging van velen, waarvan ik enkele mensen speciaal wil benoemen.

Op de eerste plaats wil ik mijn promotoren bedanken. Beste Philippe, dankzij jouw vooruitstrevende ideeën kreeg ik de kans om aan dit project te beginnen. Je nooit aflatende enthousiasme en het in mij gestelde vertrouwen hebben me altijd gemotiveerd om me verder te ontwikkelen binnen het onderzoek. Ondanks je ontzettend drukke functie was er altijd tijd en aandacht om over wetenschap te praten. Ik ben dan ook zeer blij dat ik de komende jaren nog steeds deel zal zijn van je onderzoeksteam. Bedankt voor alle mogelijkheden die me zijn geboden!

Dear Brad, it has been a pleasure working with you. Your support and guidance during this period have been very valuable to me. You were always willing to share your vision and ideas. I have enjoyed discussion science with you and hope we will do so in the future. Thank you for everything!

Dear Paul, during the last part of my thesis I got the opportunity to work with you. I thank you for the continued interest you show in all my research projects; your advice has definitely improved this thesis. You have taught me how much fun bioinformatics are and really helped me to lift my skills to the next level. Actually you are the one who inspired me to pursue a career in bioinformatics. From the first day I visited OICR you welcomed me as 'one of your own', I'll never forget that. I am proud to be an official Boutros lab member for the next couple of years. (Now don't fire me ☺).

I'd like to thank the members of the assessment committee (Prof. dr. F. Ramaekers, Prof. dr. A. Begg, Dr. C. Evelo, Prof. dr. M von Meyenfeldt and Dr. Q. Morris), for freeing time from their busy schedules to assess my PhD dissertation.

Lieve Chanti, vanaf het begin bij MAASTRO lab kunnen we het goed met elkaar vinden. In de loop van de tijd zijn we dan ook meer dan collega's van elkaar geworden. Niet te vergeten wil ik jou (en Twan) heel hartelijk bedanken voor de gastvrijheid en gezellige avondjes tijdens mijn bezoeken aan Toronto. Een tijdje terug mocht ik naast jou staan en ik ben verheugd dat jij nu naast mij wilt staan!

Lieve Sarah, mijn kantoormaatje, een paar jaar geleden begon je als student bij de mCAT groep en kwam je op de 6^{de} etage op kantoor erbij zitten. Sindsdien hebben we al heel wat gedeeld. Ik was dan ook blij dat je besloot om te blijven. Jij kunt me altijd aan het lachen maken. Dankjewel dat je mijn paranimf wilt zijn.

Natuurlijk wil ik alle (oud-)collega's van MAASTRO lab bedanken. Natasja, wat had ik zonder jou ontmoeten? Dankjewel voor alle hulp, die honderden RNA isolaties en PCR's zal ik niet vergeten! En dan Kimmie (Kim S) in de periode dat ik labwerk deed kon ik altijd bij jou terecht voor hulp met het aanleren van nieuwe technieken, bedankt! Ludwig, al bijna vanaf het begin delen we een kantoor, dankjewel voor de werkdiscussies, gezellige gesprekken en samenwerking binnen de mCAT groep. Daarnaast waren de ESTRO cursussen niet hetzelfde geweest zonder Hugo, Kasper, Chantal, Twan en Barry, dankjulliewel! Voor de administratieve ondersteuning wil ik graag Carla en Rianne hartelijk bedanken. Verder wil ik alle anderen, Kim P, Marco, Roger, Jan, Sanaz, Janneke, Arjan, Rianne, Tom, Alexey, Marc, Michael, Renaud, Nele, Asferd, Nejla, Hilda, Sherry, Younan, Mieke en Marianne bedanken voor de leuke tijd!

Ook een woord van dank aan Ramon van pulmonologie en de 'polycomb'-groep: Willen, Hanneke, Peggy, Frank en Guus, voor de discussies tijdens de labmeetings.

Off course I cannot forget to mention the Boutros lab. Although my visits to Toronto were brief I have had the pleasure to meet most of you. I really appreciated the diversity of people and the atmosphere in the lab, with everyone taking interest in the work of others. Thank you, John, Michelle, Denise, Stephenie, Syed, Daryl, Cindy, Jennifer, Ken, Dorota, Andrew, Kar, Janice, Ying, Rui, Francis, Jason, Guillermo, Yonathan, Christine, Mehrdad, Sylvia, Blair, Nickolas, Renee, Amin, Shaoming, Alex and Hanbert, for the warm welcome, fun discussions and enjoyable outings at O'Gradys. I wish you all the best and am looking forward to seeing you again soon!

I'd also like to say thanks to all the people from other research institutes whom I had the opportunity to collaborate with. I learnt from all of you!

Ook wil ik een aantal mensen binnen MAASTRO clinic bedanken voor hun belangstelling. In het speciaal wil ik de fysica groep benoemen. Door de leuke tijd bij jullie tijdens mijn stage ben ik aan dit project begonnen.

Dan wil ik graag mijn oud-studiegenoten bedanken. Ook al zien we elkaar niet vaak het is altijd gezellig. ME-ers: Vera, Anne, Cathryn, Michiel, Ellen, Ruud, Karin, Karolien, Marjon en Egon bedankt!

Alle mensen binnen de dans/gym-vereniging PHDES, in het speciaal de jazzdance groepen C en B 16+ en de jazzdance (hulp)leiding, bedankt voor de gezelligheid, leuke activiteiten en ontspanning. Heerlijk om zowel voor als in de groep te kunnen afschakelen van het werk. Nathalie, dankje voor de gezelligheid in de les en onze fijne koffieleut-afspraakjes.

Rilana, Celine, Denise, Kyara, Kim, Tessa, Noortje en Lenny met jullie heb ik altijd de grootste lol. Lekker een avondje op de dansvloer in Bolero of Valkenburg en alle stress is vergeten. Lieve vriendinnetjes, dankjulliewel!

Jolanda, Ingrid en Debby, wat ben ik blij met jullie als vriendinnen. Jullie staan altijd voor me klaar, of het nu gaat om het delen van leuk nieuws of voor een luisterend oor als ik het even niet meer zag zitten. Ik zal altijd uitkijken naar onze traditie van vier gezellige jaarlijkse etentjes. Dankjulliewel mijn BFF's, voor jullie vriendschap, steun, begrip, enthousiasme en gewoon voor wie jullie zijn.

Ook wil ik even de buurtjes, Dania, Peter, Maud en Howard bedanken voor de gezelligheid in de straat. Op naar nog vele gezellige cocktail-avondjes en bbq's!

Tot slot wil ik graag mijn familie bedanken voor hun interesse in mijn werk, steun en mooie momenten die we met elkaar hebben mogen delen. Een aantal van deze mensen wil ik graag speciaal vermelden.

Opa, oma, ome Hein en tante Liesje, ik ben heel blij dat jullie bij mijn promotie aanwezig zullen zijn. Bedankt voor jullie belangstelling in alles wat ik doe.

René en Wies, jullie huis staat altijd voor ons open. Dankjulliewel voor alle zorgen en gezelligheid in Venray. Ook Anke en Roelf wil ik graag bedanken. Ik wens jullie alle geluk met jullie kersverse gezinnetje.

Tom en Niek, ik vind het leuk dat jullie altijd zo benieuwd zijn naar mijn onderzoek. Tom heel veel succes met de specialisatie tot neuroloog. Niek je wordt zeker een fantastische Master in Music! Dankjewel dat we een jaar bij jou en Jenny hebben mogen wonen. Broertjes, ik ben trots op jullie en wens jullie een geweldige toekomst.

Lieve mam, bij jou kan ik altijd terecht. Dankjewel dat je me van jongs af aan hebt gestimuleerd om mijn dromen te volgen en altijd in me hebt geloofd. Ik had hier waarschijnlijk niet gestaan zonder jouw aanmoediging. Fijn om te weten dat er nog iemand net zo trots als jij over mijn schouder mee zal kijken vandaag. Mam dankjewel voor alles!

Roel, mijn maatje, bij jouw is het altijd fijn thuis komen. Dankjewel voor je steun, begrip en liefde de afgelopen jaren. Lief dat je altijd naar mijn verhalen over microarrays en signatures wilde luisteren. Ik kijk uit naar onze toekomst samen!

Curriculum Vitae



Curriculum Vitae

Maud Helena Wilhelmina Starmans was born on July 18th 1982 in Heerlen, the Netherlands and spend her childhood in Landgraaf. After high school at the Eijkhagen College in Landgraaf, she started the Bachelor Biomedical Engineering at the Eindhoven University of Technology in 2000. During this period she worked as an assistant teacher at the Eindhoven University of Technology and as an assistant researcher at the Department of Radiation Oncology (MAASTRO clinic) at Maastricht University. She obtained her Bachelor of Science degree in 2003 and started her Master in Medical Engineering at the same faculty. She finished her graduation work within the physics group at MAASTRO clinic and obtained her Master of Science degree in 2005. At the end of 2005, Maud started as a PhD student within MAASTRO lab studying the utility of gene expression microarrays, under the supervision of Prof. Dr. P. Lambin and Prof. Dr. B.G. Wouters. During her PhD training Maud went to Canada as a visiting scientist at the Ontario Institute for Cancer Research, where she was supervised by Dr. P.C. Boutros. She completed her PhD thesis in 2011 in the field of bioinformatics. From April 2011 she continued research in cancer-related bioinformatics as a post-doctoral fellow at the Ontario Institute of Cancer Research and MAASTRO.

List of publications

List of publications

* contributed equally

- ⌘ **Starmans M.H.W.**, Pintillie M., John T., Der S.D., Shepherd F.A., Jurisica I., Lambin P., Tsao M.S., Boutros P.C. Validation of a microarray-based marker for non-small cell lung cancer: sensitivity to data pre-processing. (*Submitted*).

- ⌘ **Starmans M.H.W.**, Lieuwes N.G., Span P.N., Haider S., Dubois L., Nguyen F., van Laarhoven H.W., Sweep F.C.G.J., Wouters B.G., Boutros P.C., Lambin P. Validation of a PCR-based proliferation signature *in vitro*, *ex vivo* and in patient studies. (*Submitted*).

- ⌘ **Starmans M.H.W.**, Chu K.C., Haider S., Nguyen F., Seigneuric R., Magagnin M.G., Koritzinsky M., Kasprzyk A., Boutros P.C., Wouters B.G.*, Lambin P*. The prognostic value of temporal *in vitro* and *in vivo* derived hypoxia gene-expression signatures in breast cancer. (*Submitted*).

- ⌘ **Starmans M.H.W.**, Fung G., Steck H., Wouters B.G., Lambin P. A simple but highly effective approach to evaluate prognostic performance of gene expression signatures. (*Submitted*).

- ⌘ Boutros P.C., Pintillie M., John T., **Starmans M.H.W.**, Der S.D., Shepherd F.A., Tsao M.S., Jurisica I. Re: Gene Expression-Based Prognostic Signatures in Lung Cancer: Ready for Clinical Use? *J Natl Cancer Inst* **102**(21): 1677-8 (2010).

- ⌘ Lambin P.*, Petit S.F.*, Aerts H.J.W.L., van Elmpt W.J.C., Oberije C.J.G., **Starmans M.H.W.**, van Stiphout R.G.P.M., van Dongen G.A.M.S., Muylle K., Flamen P., Dekker A.L.A.J., de Ruyscher D. The ESTRO Breur Lecture 2009. From population to voxel-based radiotherapy: Exploiting intra-tumour and intra-organ heterogeneity for advanced treatment of non-small cell lung cancer. *Radiother Oncol* **96**(2):145-52 (2010).

- ⌘ **Starmans MHW**, Zips D, Wouters BG, Baumann M, Lambin P: The use of a comprehensive tumour xenograft dataset to validate gene signatures relevant for radiation response. *Radiother Oncol* **92**(3):417-22 (2009).

- ⌘ **Starmans MHW**, Krishnapuram B, Steck H, Horlings H, Nuyten DSA, van de Vijver MJ, Seigneuric R, Buffa FM, Harris AL, Wouters BG, Lambin P: Robust prognostic value of a knowledge-based proliferation signature across large patient microarray studies spanning different cancer types. *Br J Cancer* **99**(11):1884-90 (2008).

- ⌘ Seigneuric R, **Starmans MHW**, Fung G, Krishnapuram B, Nuyten DS, van Erk A, Magagnin MG, Rouschop KM, Krishnan S, Rao RB, Evelo CT, Begg AC, Wouters BG, Lambin P: Impact of supervised gene signatures of early hypoxia on patient survival. *Radiother Oncol* **83**(3):374-82 (2007).

- ⌘ Duisters C, Beurskens H, Nijsten S, **Starmans M**, Wanders S, Verschueren T, Lambin P, Minken A, De Ruyscher D: Palliative chest irradiation in sitting position in patients with bulky advanced lung cancer. *Radiother Oncol* **79**(3):285-7 (2006)