Multiple Robust Signatures for Detecting Lymph Node Metastasis in Head and Neck Cancer

Paul Roepman,1 Patrick Kemmeren,1 Lodewijk F.A. Wessels,2,3 Piet J. Slootweg, and Frank C.P. Holstege1

Abstract

Genome-wide mRNA expression measurements can identify molecular signatures of cancer and are anticipated to improve patient management. Such expression profiles are currently being critically evaluated based on an apparent instability in gene composition and the limited overlap between signatures from different studies. We recently identified a primary tumor signature for detection of lymph node metastasis in head and neck squamous cell carcinomas. Before starting a large multicenter prospective validation, we have thoroughly evaluated the composition of this signature. A multiple training approach was used for validating the original set of predictive genes. Based on different combinations of training samples, multiple signatures were assessed for predictive accuracy and gene composition. The initial set of predictive genes is a subset of a larger group of 825 genes with predictive power. Many of the predictive genes are interchangeable because of a similar expression pattern across the tumor samples. The head and neck metastasis signature has a more stable gene composition than previous predictors. Exclusion of the strongest predictive genes could be compensated by raising the number of genes included in the signature. Multiple accurate predictive signatures can be designed using various subsets of predictive genes. The absence of genes with strong predictive power can be compensated by including more genes with lower predictive power. Lack of overlap between predictive signatures from different studies with the same goal may be explained by the fact that there are more predictive genes than required to design an accurate predictor. (Cancer Res 2006; 66(4): 2361-6)

Introduction

Microarray analysis has the potential to change the diagnosis and treatment of cancer (1, 2). Genome-wide gene expression measurements have been used to identify expression signatures capable of estimating a patient’s survival rate and treatment response (3, 4) and to predict the metastatic potential of primary tumors (5). Such expression profiles or signatures are expected to improve treatment strategies by providing a more personalized therapy, based for example on disease severity (6). As yet, the majority of signatures are still in a developmental stage. Prospective validation of the first profiles has been launched at institutes in Europe and the United States (2). These clinical trials are done on a large number of patients, require a great investment, and can only be carried out for profiles showing strong potential.

Despite the possible benefits, genome-wide studies for improvement of cancer diagnostics are currently being critically evaluated (7–10). Several microarray studies have identified gene sets capable of predicting a similar prognostic outcome, such as survival rate of breast cancer patients (3, 5, 11, 12). Interestingly, the overlap between the predictive gene sets from these different studies is limited to only a few genes. A recent analysis of microarray signatures found that the gene composition of expression signatures depends on the samples that were used for building the signature (7). Although the instability in gene composition is not necessarily a negative property of signatures, it does not simplify the task of choosing which genes are the best candidates for designing a diagnostic predictor.

Recently, we have identified a signature for detection of lymph node metastasis in patients with head and neck cancer based on gene expression measurements in the primary tumor (13). The potential clinical relevance of this signature resides in the difficulties for currently diagnosing the absence of lymph node metastasis in patients with head and neck cancer. Many patients receive inappropriate treatment due to difficulties in detection of metastases in the cervical lymph nodes (14, 15). The identified expression signature has the potential to improve diagnosis and treatment of head and neck cancer, particularly by reducing the number of patients given unnecessary neck surgery. The molecular signature has been validated on an independent set of tumor samples to make sure that the signature was not overfitted on the training samples and also works on new samples (13), as has been previously advocated (16). Independent validation of this signature showed an accuracy of 100% for metastasis-free predictions with an overall accuracy of 86% for all samples. Importantly, no false-negative predictions were made. Current clinical diagnosis of these patients showed an overall accuracy of 68% and included five false-negative predictions. The results of the validation set show the clinical potential of the signature. A large multicenter prospective validation study is required to confirm this potential before the signature can be applied in patient management.

Before starting such a large validation study, we decided to thoroughly evaluate the optimal gene composition of the signature (7, 17), also because the signature showed higher accuracy on samples collected later, possibly due to prolonged sample storage time (13). We report here that the initial set of predictive genes for lymph node metastasis in patients with head and neck cancer is a
subset of a larger group of 825 genes with significant predictive power. This is in agreement with earlier observations (7), and for the head and neck metastasis profile, we conclude that this is because there are many genes with a similar expression pattern across the sample collection. In contrast to other profiling studies, the predictive head and neck metastasis signature has a more stable gene composition, with a larger number of genes used in all tested predictors. Strikingly, exclusion of the most frequently occurring predictor genes could be compensated by increasing the number of genes included in the signature. Together, these analyses reveal the most comprehensive set of predictive genes that can be included in further development of a diagnostic tool for lymph node metastasis.

Material and Methods

Tumor samples and data accessibility. Head and neck squamous cell carcinoma (HNSCC) samples were processed and analyzed as described elsewhere (13). MIAME (18) compliant microarray data in microarray gene expression markup language (MAGE-ML; ref. 19) have been deposited in ArrayExpress5 with the following accession numbers: microarray layout, A-UMCU-3; tumor data, E-UMCU-11. The analysis in the study was done using 66 primary head and neck tumors that were surgically removed from the patients between 1998 and 2001. The tumor samples fulfilled the following criteria: biopsy-proven HNSCC in the oropharynx or oral cavity with no previous malignancy in the head and neck region and tumor sections containing at least 50% tumor cells. Clinical staging of the cervical lymph nodes was done according to the Netherlands national consensus guidelines for oral cavity and oropharynx, by palpation of the neck region, followed by bilateral sonography examination, computed tomography, and/or magnetic resonance imaging. Suspected nodes were subjected to aspiration cytology. Based on the clinical staging of the neck, 38 patients were classified as N0 (metastasis-free) and 28 as N+ (presence of metastasis in lymph nodes). Patients assessed to be N0 underwent selective dissection (levels I-III, supra-omohyoidal neck dissection), and patients assessed to be N+ underwent comprehensive dissection (levels I-V, radical neck dissection; ref. 14). In agreement with previous studies (20, 21), there was a high prevalence of smoking within the patient cohort. The samples from the three nonsmokers did not behave discordantly with regard to clinical assessment, microarray prediction, or histologic determination of N status, although we note that this group size is too small to result in statistically meaningful analyses.

Supervised classification. To remove the possible negative influence of the older tumor samples that were surgically removed in 1996 and 1997, we built a new molecular signature for prediction of lymph node metastasis. The supervised classification procedure was identical to the one used previously (13). We left out the 38 tumor samples from 1996 and 1997 and combined the initial training and test sets into a new training set containing 66 tumor samples from 1998 to 2001. After preprocessing the expression data of the 21,329 genes on the microarray, 3,064 were found to be differentially expressed \( (P < 0.01) \) in at least 15 of the 66 tumor samples. These 3,064 genes were used for designing the predictor with the highest overall accuracy as described previously (13). Briefly, the set of samples were iteratively divided into training (two thirds) and test (one third) sets. On the training set, using a 10-fold cross-validation procedure, the optimal set of genes to employ in the classifier was determined based on the signal-to-noise ratio and classification performance. Performance of this optimal set of genes was validated on the one-third test set. This 3-fold cross-validation loop was repeated 100 times to select the final list of predictive genes used within the molecular signature.

Multiple training approach. A multiple training approach similar to the one used by Michiels et al. was used to study the stability of the identified signature based on the 66 tumor samples from 1998 to 2001. The tumor samples were randomly divided into a training set and test set using a 10-fold cross-validation procedure. Based on the training set, \( P \)s were calculated for all 3,064 differentially expressed genes based on the difference in expression between \( N^- \) and \( N_0 \) tumor samples (Student’s \( t \) test). The set of genes with the lowest \( P \)s (i.e., most predictive) was used for prediction of the test samples by calculating the correlation with the average \( N^- \) and average \( N_0 \) training profile and, based on these correlations, classifying the test samples as \( N_0 \) or \( N^- \). Repeating this resampling procedure a thousand times resulted in multiple predictions for each tumor sample, based on the different predictive gene sets.

Signature composition analysis. The multiple training approach was done for sets of 50, 100, and 200 genes, which were used for building predictive signatures. Investigation of the stability in signature gene composition was done by scoring each gene for the number of times it was included in a predictive signature. The selection ranged from 0% (used in none of the signatures) to 100% (used in all thousand generated signatures). The complete set of predictive genes was defined as those genes that were selected at least once during the repeated sampling of the multiple training approach, whereby either 50, 100, or 200 genes were selected. The predictive set of 825 genes are found upon repeated sampling of signatures constructed of 200 genes.

Results

The recently reported signature for detection of lymph node metastasis in patients with head and neck cancer showed a

---

5 http://www.ebi.ac.uk/arrayexpress/.
strong predictive performance with an independent validation set. The accuracy for the oldest samples in the training set was lower, perhaps due to prolonged storage of these samples (13). To investigate the influence of the older samples on the composition and performance of the signature, we left out the oldest samples and rebuilt the signature using 66 samples from 1998 to 2001 (44 from the initial training set and 22 from the validation set). This signature was designed in exactly the same way as the previous published signature (Materials and Methods). Importantly, the predictive outcome of the signature on the newer samples is similar to the original (85% accuracy), indicating that the previous presence of older samples did not interfere with the performance on the newer samples. Interestingly, the overlap in predictive genes found in both predictors is limited to 49 genes (Fig. 1A). Cursory examination of the signature genes indicated that the incomplete overlap is due to the presence of a large number of genes with similar patterns of expression across the samples (e.g., Fig. 1B). This indicates that many predictive genes can be interchanged without influencing the predictive outcome and suggests that multiple, different gene sets can be made that are useful for accurate prediction. Because the goal of this work is to detect the most useful set of genes for head and neck metastasis prediction, we decided to investigate this further.

To study whether different gene sets show similar predictive outcome, we used a multiple training approach similar to the one Michiels et al. used for validating prognostic significance of previously published microarray signatures (7). Samples were randomly divided into training and test sets using a 10-fold cross-validation procedure. The 50, 100, or 200 most predictive genes were selected and used to classify the metastasis status of the test samples. Repeating this procedure generated 3,000 different predictive gene sets consisting of 50, 100, or 200 genes. Although the sets had a different gene composition, the power to discriminate between histologically determined metastasis (N+) and metastasis-free (N0) tumors remained similar. The predictive outcome on individual tumor samples was generally similar, with decreased variance for larger gene sets (100 and 200 genes; Fig. 2A-C).

The similar predictive outcome of the multiple gene sets is not caused by a fixed set of genes present in all signatures. In the multiple signatures consisting of 50, 100, or 200 genes, 10, 27, or 49 genes were always selected respectively, and 41, 88, and 180 genes were selected in at least half of each of the thousand signatures (Fig. 3A-B). These frequently selected genes account for only 5% of the total of 825 predictive genes selected at least once during the multiple sampling approach (Supplementary Table S1). This degree of stability is higher than for the two most stable signatures previously analyzed by Michiels et al. (7). The hepatocellular carcinoma predictive signature of lizuka et al. (22) showed 13 genes selected in at least half of the signatures with none of these genes selected always (Fig. 3C). The breast cancer data set of van’t Veer et al. (3) showed 24 genes selected in at least 50% of the signature with one gene selected always (Fig. 3D).

Genes commonly used in the multiple training signatures show a strong overlap with the predictive genes identified using the initial two-step supervised classification approach on the same 66 samples. Eighty-three percent of the genes present in the majority of the multiple signatures were also identified using the two-step supervised classification method (Fig. 3B, gray columns). In comparison, the overlap in gene selection between the multiple training approach and the originally published signature was 58% in the van’t Veer study and 38% in the lizuka data set (Fig. 3C-D, gray columns), which represent the studies with the highest stability as analyzed by Michiels et al. (7).

Finding genes that are used in the majority of accurate signatures indicates that these genes are important to include in any signature for head and neck metastasis. To test whether these frequently selected genes were pivotal for accurate prediction, the number of genes that are also present in the originally identified signatures.

Figure 2. The predictive outcome of different signatures is stable. Predictive correlation outcome of 66 tumors samples using a multiple training approach. A thousand different molecular signatures of 50 (A), 100 (B), or 200 (C) genes were used to predict each sample approximately 100 times. Blue, samples from patients without metastasis; red, samples from patients with lymph node metastasis. Shaded area, 95% confidence interval for the sample predictions.

Figure 3. Stability of signature composition. A, table containing the number of predictive genes selected using the multiple training approach. Numbers are shown for genes selected always, at least 50% and at least once in the predictive lists of 50, 100, and 200 genes, respectively. B, 287 genes were used to design all various 50-gene signatures, 489 genes for the 100-gene signatures, and 825 genes for the 200-gene signatures. (B-D) Genes selected in at least 50% of the predictive lists of 50 genes in the head and neck data set (B), the signature of lizuka et al. (C), and the predictive signature of van’t Veer et al. (D). * genes selected in all signatures. Gray columns, genes that are also present in the originally identified signatures.
the 825 predictive genes that were selected at least once during the repeated sampling procedure were ordered according to the frequency of selection and divided into subsequent sets of 50 genes by applying a moving window with steps of 25 genes (i.e., 1-50, 25-76, 51-100, etc.; Fig. 4A, bottom). These subsequent sets were used for classification of the tumor samples. The predictive accuracy decreases for sets containing less frequently selected genes but does not drop considerably below the current clinical accuracy of 75% (Fig. 4A). Signatures without the frequently selected genes still show predictive power. This indicates that the frequently selected genes are not essential for prediction, but that they do contribute more towards improved accuracy. Strikingly, the observed decrease in predictive accuracy can be completely compensated by increasing the number of genes used in a signature (Fig. 4B). For enlarged signatures of less frequently selected genes, the accuracy remains around 86%, similar to the accuracy of the original predictor. In other words, increasing the quantity of the predictive genes can compensate for reduced quality. Signatures built from large random sets of 100 to 200 predictive genes resulted in a stable predictive outcome with an accuracy of 80% to 90% (Fig. 5A-E). In conclusion, this indicates that numerous combinations of predictive genes can be used for accurate prediction. In total, we have identified a large set of 825 predictive genes from which multiple accurate predictive signatures can be derived (Fig. 6).

**Discussion**

We report here that our initially identified set of predictive genes for detection of lymph node metastasis in patients with head and neck cancer (13) is a subset of a larger group of predictive genes. Using a resampling approach, we have identified a large set of 825 genes that can be used for prediction of metastasis. Based on this group of genes, multiple predictive signatures can be made with high predictive accuracy. The phenomenon that different sets of genes can be used for accurate prediction is not exclusive for this study but is becoming apparent in other cancer profiling studies (7, 17). Due to minor differences in gene expression, different genes are selected for optimal prediction when the signature is built using different samples, especially when comparing studies that have been done in different institutes (3, 12). This instability in gene composition of different predictive signatures is not detrimental as long as the predictive outcome and accuracy remain similar. Different gene sets can give comparable results because individual genes that show equal expression patterns can be interchanged without affecting the signature profile and the predictive outcome.

Although the predictive signature for lymph node metastasis shows instability in gene composition, it is more stable than other molecular profiles analyzed similarly by Michiels et al. (7). A possible explanation for this higher stability is the reduction of biological variation by analyzing tumors from only two locations.
within the head and neck region: oropharynx and oral cavity. Another possible explanation of the increased stability is related to the complexity of the different disease characteristics considered in the different studies. The head and neck signature predicts the presence of metastasis in lymph nodes that are close to the site of the primary tumor. Predicting a more complex or long-term patient trait, such as survival rate and development of distant metastasis (3, 12), likely depends on more factors and developmental pathways (23). Therefore, prediction of more complex characteristics over time is probably susceptible to more variation, perhaps resulting in a less stable predictive signature.

Predictive signatures lacking the most frequently selected genes remains reasonably accurate, a phenomenon that was also found by Ein-dor et al. (17) when reanalyzing the breast cancer profile by van’t Veer. In addition, here, we show that the reduction in predictive power can be fully compensated by increasing the number of genes used in the signatures. This implies that for expression signatures both the quality and the quantity of the genes are important for predictive accuracy. Selection of the most frequently selected genes is nevertheless helpful for reducing the number of genes in a signature.

Due to the interchangeability of predictive genes, there is no single set of genes with optimal predictive accuracy. Various signatures can be identified by different institutes or simply by using different samples, and the identified gene sets with optimal predictive accuracy will differ due to minor differences in the analyzed samples. This does not mean that the different signatures are based on random noise in the data sets, as Michiels et al. concluded (7). Although the genes identified as most predictive can differ between different studies, the overall predictive profiles can be similar, resulting in an identical predictive outcome.

Now that we know that none of the head and neck lymph node metastasis predictive genes is essential for accurate prediction, is it wise to try to make a predictive list as small as possible? A molecular signature that is based on more genes is likely to be less prone to biases towards specific samples. When certain genes within a larger signature show lower predictive power for new samples, other predictive genes in the signature may compensate this effect. Ma et al. recently identified a set of only two genes that could accurately predict tamoxifen treatment outcome in breast cancer patients (4). When Reid et al. tried to validate this two-gene signature on independent samples, they were unable to show predictive power of these two genes (8). This example clearly illustrates the risk of reducing a signature to a small number of genes without a thorough validation on independent samples.

The set of lymph node metastasis predictive genes reported here also sheds light on the development of metastasis. Two interesting overrepresented functional categories within the set of predictive genes are binding to the extracellular matrix and protease activity for degradation of the extracellular matrix (Supplementary Fig. S2). Both categories are up-regulated in tumors that metastasize to the lymph nodes. These two categories seem contradictory; however, they support the theory that tumor cells gain mobility by an interplay between anchoring to the extracellular matrix and degradation of this matrix (24). In this way, groups of tumor cells can move through the surrounding tissue by degrading the extracellular matrix while retaining cell to cell and cell to extracellular matrix contact. The invasion in the surrounding tissue is not solely caused by the tumor cells and the extracellular matrix but also includes nontumor cells in the tumor microenvironment, such as stromal fibroblasts, lymphocytes, and macrophages (reviewed in refs. 25, 26).
new diagnostics to identify tumors with metastatic potential should therefore not exclusively focus on processes in the tumor cells but also include the tumor microenvironment (27). Targeting both the tumor and nontumor cells may therefore offer a more efficient way to diagnose and treat cancer.

References
Multiple Robust Signatures for Detecting Lymph Node Metastasis in Head and Neck Cancer

Paul Roepman, Patrick Kemmeren, Lodewijk F.A. Wessels, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/66/4/2361

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2006/02/14/66.4.2361.DC1

Cited articles
This article cites 27 articles, 6 of which you can access for free at:
http://cancerres.aacrjournals.org/content/66/4/2361.full.html#ref-list-1

Citing articles
This article has been cited by 6 HighWire-hosted articles. Access the articles at:
/content/66/4/2361.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.