Molecular Staging for Survival Prediction of Colorectal Cancer Patients

STEVEN ESCHRICH¹, IVANA YANG², GREG BLOOM¹, KA YIN KWONG², DAVID BOULWARE¹, ALAN CANTOR¹, DOMENICO COPPOLA¹, MOGENS KRUHØFFER³, LAURI AALTONEN³, TORBEN F. ORNTOFT³, JOHN QUACKENBUSH²,⁴, TIMOTHY J. YEATMAN¹

¹The H. Lee Moffitt Cancer Center and Research Institute, Departments of Surgery, Pathology, and Biostatistics, Tampa, Florida, USA
²The Institute for Genomic Research, Rockville, Maryland, USA
³The Molecular Diagnostic Laboratory, Dept. Of Clinical Biochemistry, Aarhus University Hospital, Skejby Denmark, Department of Medical Genetics, Biomedicum in Helsinki, Finland.
⁴Department of Biochemistry, The George Washington University, Washington, DC, USA and Department of Statistics, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD, USA

Running Title: Molecular Staging for Survival Prediction

Address reprint requests to Timothy J. Yeatman, M.D at the H. Lee Moffitt Cancer Center and Research Institute, 12902 Magnolia Drive, SRB-2, Tampa, FL 33612, telephone 813-979-7292, fax 813-632-1433, or at yeatman@moffitt.usf.edu or to johnq@tigr.org.

-Supported by grants from the National Cancer Institute.
-This manuscript contains original material.
-The human investigations were performed after approval by the University of South Florida IRB, and in accord with an assurance filed with and approved by the Department of Health and Human Services. A waiver of informed consent was filed.
ABSTRACT

PURPOSE: The Dukes’ staging system is the gold standard for predicting colorectal cancer prognosis; however, accurate classification of intermediate-stage cases is problematic. We hypothesized that molecular fingerprints could provide more accurate staging and potentially assist in directing adjuvant therapy. METHODS: A 32k cDNA microarray was used to evaluate 78 human colon cancer specimens and these results were correlated with survival. Molecular classifiers were produced to predict outcome. RESULTS: Molecular staging, based on 43 core genes, was 90% accurate (93% sensitivity, 84% specificity) in predicting 36 month overall survival in 78 patients. This result was significantly better than Dukes’ staging (P=0.03878), discriminated patients into significantly different groups by survival time (P<0.001, log rank test), and was significantly different from chance (P<0.001, 1000 permutations). Furthermore, the classifier was able to discriminate a survival difference in an independent test set from Denmark. Molecular staging identifies patient prognosis (as represented by 36 month survival) more accurately than the traditional clinical staging, particularly for intermediate Dukes’ stage B and C patients. The classifier was based on a core set of 43 genes including osteopontin and neuregulin, which have biological significance for this disease. CONCLUSION: These data support further evaluation of molecular staging to discriminate good from poor prognosis patients, with the potential to direct adjuvant therapy.

INTRODUCTION

Colorectal cancer staging is currently based solely on simple clinicopathologic features such as bowel wall penetration and lymph node metastasis. Unfortunately, clinical staging systems often fail to discriminate the biological behavior of a large number of tumors, resulting in the systematic over-treatment or under-treatment of patients with adjuvant therapies.
over 70 years ago\textsuperscript{1}, the now modified Dukes’ staging system provides adequate prognostic information for patients staged as A or D. However, the intermediate stages B and C are not very useful in discriminating good from poor prognosis cases. Additionally, application of this staging system results in the potential over-treatment or under-treatment of a significant number of patients and it can only be applied after complete surgical resection rather than after a pre-surgical biopsy. Recently developed microarray technology has permitted development of multi-organ cancer classifiers\textsuperscript{2,3}, identification of tumor subclasses\textsuperscript{4-6}, discovery of progression markers\textsuperscript{7,8} and prediction of disease outcome in many types of cancer\textsuperscript{9-11}. Unlike clinicopathologic staging, molecular staging has promise in predicting the long-term outcome of any one individual based on the gene expression profile of the tumor at diagnosis. Inherent to this approach is the hypothesis that every tumor contains informative gene expression signatures that, at the time of diagnosis, can direct the biological behaviour of the tumor over time\textsuperscript{12}.

**METHODS**

**Tumor Samples (Moffitt Cancer Center)**

We developed a colorectal cancer survival classifier using 78 tumor samples including 3 adenomas and 75 cancers. Informative frozen colorectal cancer samples were selected from the Moffitt Cancer Center Tumor Bank based on evidence for good (survival $> 36$ month) or poor (survival $< 36$ month) prognosis from the Tumor Registry. The samples were initially selected such that all patients had follow-up for at least 36 months. Forty-eight samples were poor prognosis and 30 were good prognosis cases. Dukes’ stages included a mixture of B, C, and D cases. Dukes’ stage A cases are very rare and were not available for analysis, however we selected 3 adenomas as examples of very good prognosis cases. Survival was measured as last
contact minus collection date for living patients, or date of death minus collection date for patients who had died. The median follow-up time was 27.9 months (range, 0.49 to 119 months), the median follow-up time among patients alive at last follow-up (26 cases) was 64 months.

Samples were microdissected (>80% tumor cells) by frozen section guidance and RNA was extracted using Trizol followed by secondary purification on RNAEasy columns. The samples were profiled on TIGR’s 32,488-element spotted cDNA arrays, containing 31,872 human cDNAs representing 30,849 distinct transcripts – 23,936 unique TIGR TCs and 6,913 ESTs, 10 exogenous controls printed 36 times, and 4 negative controls printed 36-72 times.

**SAM Survival Analysis**

The first analysis of the colorectal cancer microarray data used the Significance Analysis of Microarrays (SAM) with censored survival data. SAM identifies genes most closely correlated with survival time and uses permutation analysis to estimate the false discovery rate (FDR). Mean-centered gene expression vectors were then clustered and visualized using Cluster 3.0 and Java TreeView 1.03. We used the two clusters of SAM-selected genes as a survival grouping and constructed Kaplan-Meier curves for these two groups. The samples were also manually grouped by Dukes’ stage for a comparison of survival times.

**Classifier Construction and Evaluation**

A leave-one-out cross-validation (LOOCV) technique was used for constructing and validating a neural network-based classifier. The samples were classified as having “good” or “poor” prognosis based on survival for more or less than 36 months, respectively. Using the leave-one-out cross-validation approach also provides the ability to rank the selected genes. The number of times a particular gene was chosen can be an indicator of the usefulness of that gene for general classification and may imply biological significance. Therefore, we examined the
genes that were consistently selected by the t-test. We focused on 43 core genes identified in 75% of the LOOCV iterations\textsuperscript{15}.

The molecular classifier was composed of two distinct steps: gene selection using a t-test and classification using a neural network. Both steps were taken after the test sample was left out (from the leave-one-out cross-validation) to avoid bias from the gene selection step. We used the top 50 genes as ranked by absolute value of the t statistic using a t test, for each cross-validation step. A feed-forward back-propagation neural network\textsuperscript{16} with a single hidden layer of 10 units, learning rate of 0.05, and momentum of 0.2 was constructed. Training occurred for a maximum of 500 epochs or until a zero misclassification error was achieved on the training set. Our experiences indicate neural networks are extremely robust to both the number of genes selected and the level of noise in these genes. We have successfully used this classifier in earlier multi-platform gene expression classification experiments\textsuperscript{3}.

\textbf{Statistical Significance}

The log-rank test was used to measure the difference in Kaplan-Meier curves, both for the initial survival analysis and for the classifier results. The classifier splits the examples into two groups: those predicted as good or poor prognosis. Classifier accuracy was reported both as overall accuracy and as specificity/sensitivity. A McNemar’s Chi-Square test was used to compare the molecular classifier with the use of a Dukes’ staging classifier. Finally, 1,000 permutations of the dataset were used to measure the significance of the classifier results as compared to chance.
RESULTS

Identification of Prognosis-Related Genes

We used SAM survival analysis to identify a set of genes most correlated with censored survival time. A set of 53 genes was found, corresponding to a median expected false discovery rate (FDR) of 28%. These genes are listed in Supplemental Table 1 and include several we believe to be biologically significant, such as osteopontin and neuregulin (see Discussion).

Figure 1A presents a graphical representation of these 53 SAM-selected genes as a clustered heat map. The figure uses only the Dukes’ stage B and C cases, whose outcome Dukes’ staging predicts poorly (see Fig 1 D). Since we clustered using only genes correlated with survival, the clusters should correspond to very different prognosis groups. The SAM-selected genes are also arranged by annotated Dukes’ stage (B and C stages only) in Figure 1B which shows little discrimination based on gene expression.

Figure 1C shows the Kaplan-Meier plot for the two dominant gene expression clusters of stage B and C samples. Clearly, these 53 genes separate the cases into two distinct clusters of patients with good prognosis (cluster 2) and poor prognosis (cluster 1) (P<0.001 using a log rank test) as expected since the genes were selected using SAM. Figure 1D presents a Kaplan-Meier plot of the survival times of Dukes’ stage B and C tumors grouped by stage, showing no statistically significant difference, demonstrating the weak potential for discrimination of these cases by Dukes’ staging. Here we demonstrate that gene expression profiles can separate good and poor prognosis cases better than Dukes’ staging. This suggests that a gene-expression based classifier could be more accurate at predicting patient prognosis than the traditional Dukes’ staging.
Performance of a 43-gene cDNA-based Colorectal Cancer Survival Classifier

A leave-one cross-validation (LOOCV) was used to evaluate a 43-gene cDNA classifier in predicting prognosis for each patient at 36 months of follow-up. The classifier accuracy was 90% (93% sensitivity/84% specificity). A log-rank test of the two predicted groups (good and poor prognosis) is significant (P<0.001), demonstrating the ability of the classifier to distinguish the two outcomes (see Fig 1E). Good prognosis was defined by survival > 36 mo and poor prognosis was defined by survival < 36 mo. Permutation analysis demonstrates the result is better than possible by chance (P<0.001 – 1000 permutations). This result is also significantly higher than that observed using Dukes’ staging as a classifier for the same group of patients (P=0.03878). The results for molecular staging are summarized in Table 1. Molecular staging identified the good prognosis cases (the “default” classification using Dukes’ staging), but also identified poor prognosis cases with a high degree of accuracy. Table 1 also shows the detailed confusion matrix for all samples in the dataset, showing the equivalent misclassification rate of both good and poor prognosis groups by our classifier. Supplemental Table 2 lists all 43 genes selected by the LOOCV approach via t-test and used to construct the cDNA classifier.

Evaluation of an Independent Test Set from Denmark

In order to further validate our classifier, we identified an independent Danish colon cancer data set, comprised of Dukes Stage B and C patients, and produced on a U133A Affymetrix GeneChip oligonucleotide-based platform. Because this data set was produced on an oligonucleotide platform instead of our cDNA platform, we first translated our gene signature into available Affymetrix probe sets using the Resourcerer program (WWW.TIGR.org). This translation reduced the gene signature from 43 genes to only 26 unique genes. Therefore, we
limited the original cDNA classifier to only genes represented on the U133A platform. Using this restricted gene signature derived solely from the Moffitt data set, we found 60 corresponding probe sets on the Affymetrix U133A GeneChip and used these genes to evaluate the survivorship of 95 Danish patients. With this approach, hierarchical clustering was used to find the most significant groups in the data. The survival of these groups was then displayed using Kaplan-Meier survival analysis. Figure 2A shows that the 26 selected genes were able to discriminate good from poor prognosis patients, despite the restrictions imposed by cross-platform analyses. Dukes’ staging was incapable of discriminating survivorship in these same patients (see Fig 2B). When applied to the Dukes B and Dukes’ C cases separately, the 26 gene signature was capable of discriminating good from poor prognosis subpopulations within each stage (see Fig 2C).

**DISCUSSION**

The benefit of adjuvant chemotherapy for colorectal cancer appears limited to patients with Dukes stage C disease where the cancer has metastasized to lymph nodes at the time of diagnosis. For this reason, the clinicopathological Dukes’ staging system is critical for determining how adjuvant therapy is administered. Unfortunately, it is not very accurate in predicting overall survival and thus its application likely results in the treatment of a large number of patients to benefit an unknown few. Alternatively, there are probably a number of patients who might benefit from therapy that do not receive it.

Molecular staging may provide more accurate predictions of patient outcome than is currently possible with clinical staging, which may, in fact, misclassify patients. Using a SAM selected set of genes derived from a genome wide analysis of gene expression, we were first able to cluster groups of patients with good and bad prognoses, suggesting that outcome-rich information was likely present in this gene expression dataset. Subsequently, a supervised
learning analysis identified a core set of 43 informative genes that appeared in 75% of the cross validation iterations and accurately predicted colorectal cancer survival. This core set was derived from a 32,000-element cDNA microarray that included both named and unnamed genes. This gene set was highly accurate in predicting survival when compared with Dukes’ staging data from the same patients.

Although Dukes’ staging works well for very good and very poor prognosis patients (Dukes’ stage A and D), currently it is not very informative when predicting long-term outcomes of intermediate prognosis patients (Dukes’ B and C), yet is the primary means for determining the administration of potentially toxic adjuvant chemotherapy. We hypothesized that molecular staging might be able to identify those Dukes’ stage B and C cases for which chemotherapy may be beneficial. The production of a cDNA classifier for survival is a first step in the validation process for molecular staging.

With our approach, we were able to determine which genes appeared to be most useful in the classifier based on their frequency of appearance in the classification set. Of these, at least two, osteopontin and neuregulin, have reported biological significance in the context of colorectal cancer. Osteopontin, a secreted glycoprotein and ligand for CD44 and αvβ3 appears to have a number of biological functions associated with cellular adhesion, invasion, angiogenesis and apoptosis. Using an oligonucleotide microarray platform, we recently identified osteopontin as a gene whose expression was strongly associated with colorectal cancer stage progression. Moreover, we have recently identified INSIG-2, one of the 43 core classifier genes, as an osteopontin signature gene (data not shown), suggesting that an osteopontin pathway may be prominent in regulating colon cancer survival. Similarly, neuregulin, a ligand for ERBB receptors, may have biological significance in the context of colorectal cancer where
current data suggest a strong relationship between colon cancer growth and the ERBB family of receptors\textsuperscript{19}. Neuregulin was recently identified as a prognostic gene whose expression correlated with bladder cancer recurrence\textsuperscript{15}.

While cross-platform analysis is challenging due to the paucity of available human data sets, the discrepancies in probes used on competitive platforms and due to the differential performance characteristics of these probes, we demonstrated that the genes selected for our cDNA-based classifier were effective in discriminating good from poor prognosis patients using a completely independent data set produced on a Danish population using an oligonucleotide platform (Affymetrix, U133A). These data provide confirmation, under the most rigorous conditions, that there is prognostic value for the identified gene signature.

Of interest is a recent report of a gene expression profile, using the U133A Affymetrix platform, to predict recurrence for Dukes’ B patients\textsuperscript{20}. The reported 23-gene signature with a performance accuracy of 78% shares no genes in common with the cDNA classifier we have produced. The absence of concordant genes could be related to many different issues including differences in the microarray platform, the samples selected for analysis, and the analytic tools used to generate the gene signatures, suggesting the need for extensive validation of any promising signature prior to clinical implementation of any gene expression signature.

We have produced the first colorectal cancer molecular staging classifier, based on the analysis of all colon cancer stages, for which accuracy exceeds that of Dukes’ staging when used to estimate prognosis on the same patients. These results suggest a molecular-based method for the discrimination of outcome for intermediate Dukes’ stages B and C, where prognosis is currently problematic, may be effective. Our classifier is based on a core set of 43 genes that
appear to have biological significance for human colorectal cancer progression. The data provided support more extensive validation of this prognostic gene signature.

**ACKNOWLEDGEMENTS**

Additional support was provided by the Biostatistics Core and the Microarray Core at the H. Lee Moffitt Cancer Center and the services provided by the Research Computing Core, University of South Florida.

**REFERENCES**

### Table 1. Results from cDNA Classifier.

LOOCV Accuracy of Molecular Staging for all tumors.

<table>
<thead>
<tr>
<th>Classification Method</th>
<th>Total Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular Staging</td>
<td>90%</td>
<td>93%</td>
<td>84%</td>
</tr>
</tbody>
</table>

Breakdown of Molecular Staging Accuracy by Dukes’ Stage.

<table>
<thead>
<tr>
<th>Dukes’ Stage</th>
<th>Molecular Staging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenoma</td>
<td>100%</td>
</tr>
<tr>
<td>B</td>
<td>87%</td>
</tr>
<tr>
<td>C</td>
<td>91%</td>
</tr>
<tr>
<td>D</td>
<td>90%</td>
</tr>
</tbody>
</table>

Confusion Matrix of cDNA Classifier Results.

<table>
<thead>
<tr>
<th>Observed/Predicted</th>
<th>Poor</th>
<th>Good</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>43</td>
<td>3</td>
<td>46</td>
</tr>
<tr>
<td>Good</td>
<td>5</td>
<td>27</td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>30</td>
<td>78</td>
</tr>
</tbody>
</table>
**Figure 1.** Training Set Details show the potential of molecular staging. Using Dukes’ stage B and C tumor samples: A) Cluster analysis of 53 SAM selected genes was performed. Red color represents over-expressed genes relative to green, under-expressed genes. The data suggest that genes can be identified that discriminate good from poor prognosis; B) Cluster analysis of SAM selected genes, grouped by Dukes’ stage B and C does not demonstrate a discriminating pattern; C) Survival curves corresponding to gene clusters; D) Survival curves for Dukes’ B and C patients; E) Survival curves for the Molecular Classifier using 78 samples tested by LOOCV.

**Figure 2.** Independent Test Set Evaluation (Denmark Test Set) using the U133A data set. A) Survival curves generated using probe sets corresponding to 26 of the Molecular Classifier genes. Using these translated probe sets, 95 tumors were clustered and censored survivorship was evaluated (P < 0.001). B) Survival curves using Dukes’ staging criteria show no significant difference in outcome. C) Survival curves grouped by both Dukes’ stage and molecular signature shows that both Dukes’ B and C cases can be further subdivided into good and poor prognosis groups.
Molecular Staging for Survival Prediction

Figure 1

A  SAM-Selected Genes
Cluster 1 (n=23)  Cluster 2 (n=22)

B  Dukes’ Grouping

C  SAM-Selected Survival Curve
Survival by Cluster (in Months)

D  Dukes’ Grouping Survival Curve
Survival by Dukes’ Stage (in Months)
Molecular Staging for Survival Prediction

Survival by Classification
cDNA Classifier LOOCV

- Good Prognosis (n=50)
- Poor Prognosis (n=48)

P < 0.001
Log Rank Test

Survival Distribution

Month

0 20 40 60 80 100 120
Figure 2

A

Denmark Survival Data

Good Prognosis (n=63)
P<0.001
Log Rank Test

Poor Prognosis (n=32)

Survival (months)

B

Denmark Survival Data

Dukes' B (52)
P=0.118
Log Rank Test

Dukes' C (63)

Survival (months)

C

Denmark Survival Data

Dukes' B Poor Prognosis (n=8)

Dukes' B Good Prognosis (n=24)

Dukes' C Poor Prognosis (n=24)

Dukes' C Good Prognosis (n=39)