### Abstract

**Purpose:** To address the progression, metastasis, and clinical heterogeneity of renal cell cancer (RCC).

**Experimental Design:** Transcriptional profiling with oligonucleotide microarrays (22,283 genes) was done on 49 RCC tumors, 20 non-RCC renal tumors, and 23 normal kidney samples. Samples were clustered based on gene expression profiles and specific gene sets for each renal tumor type were identified. Gene expression was correlated to disease progression and metastasis gene signature was derived.

**Results:** Gene signatures were identified for each tumor type with 100% accuracy. Differentially expressed genes during early tumor formation and tumor progression to metastatic RCC were found. Subsets of these genes code for secreted proteins and membrane receptors and are both potential therapeutic or diagnostic targets. A gene pattern ("metastatic signature") derived from primary tumor was very accurate in classifying tumors with and without metastases at the time of surgery. A previously described "global" metastatic signature derived by another group from various non-RCC tumors was validated in RCC.

**Conclusion:** Unlike previous studies, we describe highly accurate and externally validated gene signatures for RCC subtypes and other renal tumors. Interestingly, the gene expression of primary tumors provides us information about the metastatic status in the respective patients and has the potential, if prospectively validated, to enrich the armamentarium of diagnostic tests in RCC. We validated in RCC, for the first time, a previously described metastatic signature and further showed the feasibility of applying a gene signature across different microarray platforms. Transcriptional profiling allows a better appreciation of the molecular and clinical heterogeneity in RCC.

Renal cell cancer (RCC) has a heterogeneous clinical presentation with up to 30% metastatic cases at initial diagnosis and ~30% of initially organ-confined cases developing metastases during follow-up at variable intervals (1, 2). Although surgery is highly effective for the treatment of localized RCC, the treatment options available for patients with metastatic disease are very limited (3, 4) and do not take into account the underlying molecular differences among different histologic RCC subtypes. Although the response rates to currently available systemic therapies for the treatment of locally advanced and metastatic disease are in general poor, there is a subset of patients with a well-documented long remission (3). The variability in clinical outcome is possibly attributable to the molecular heterogeneity of RCC, which has not yet been fully elucidated.

Transcriptional profiling has emerged as a powerful approach to identify new cancer classes (class discovery), assigning tumors to known classes (class prediction; refs. 5, 6) and predicting clinical outcome solely based on gene expression (7), and has the potential to affect patient diagnosis, staging, prognosis, and treatment. Differentially expressed genes between RCC and normal kidney tissue as well as between RCC samples associated with different outcomes have been described in literature (8–10).

In this large study, we addressed additional clinical and biological aspects of RCC, focusing on disease progression and metastasis. We investigated the gene expression profile of our clear cell RCC samples with respect to tumor progression, identified a clear cell RCC metastatic gene signature, validated these gene signatures in independent external data sets, and successfully applied a previously described "global" metastatic gene signature to our clear cell RCC samples. We also identified

<table>
<thead>
<tr>
<th>Authors' Affiliations:</th>
<th>Purpose: To address the progression, metastasis, and clinical heterogeneity of renal cell cancer (RCC).</th>
</tr>
</thead>
<tbody>
<tr>
<td>1Beth Israel Deaconess Medical Center Genomics Center and Dana-Farber/Harvard Cancer Center Proteomics Core, 2Bioinformatics Core of Beth Israel Deaconess Medical Center Genomics Center and Harvard Medical School, and 3Division of Hematology/Oncology, Beth Israel Deaconess Medical Center, Boston, Massachusetts; 4Department of Urology and Pediatric Urology and 5Institute of Pathology, Johann Wolfgang Goethe University, Frankfurt, Germany; and 6University of California-Los Angeles Kidney Cancer Program, Department of Urology, Jonsson Comprehensive Cancer Center, and David Geffen School of Medicine at University of California-Los Angeles, Los Angeles, California</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Supplementary data for this article are available online at http://www.bidgenomics.org/KidneyCancer/index.html.

**Requests for reprints:** Towia A. Libermann, Beth Israel Deaconess Medical Center Genomics Center and Harvard Medical School, Harvard Institutes of Medicine, 4 Blackfan Circle, Boston, MA 02115. Phone: 617-667-3393; Fax: 617-975-5299; E-mail: tliberma@bidmc.harvard.edu.

© 2005 American Association for Cancer Research.

doi:10.1158/1078-0432.CCR-04-2225
gene signatures that were able to accurately characterize RCC subtypes and other renal tumors, such as oncocytomas and transitional cell cancers of the renal pelvis (TCC), both of which are important in the differential diagnosis of RCC and were underrepresented in previous studies.

Materials and Methods

Tumor samples. A total of 65 frozen primary tumors (23 clear cell RCC, 13 papillary RCC, 7 chromophobe RCC, 10 TCC, and 12 oncocytomas), 24 frozen normal kidney tissue samples, and 10 frozen metastatic tumors (9 distant and 1 nodal metastatic clear cell RCC) were received from the Department of Urology, Johann Wolfgang Goethe University (Frankfurt, Germany), Ardais and Collaborative Genomics and processed with approval from the Institutional Review Board/ Human Subject Research Committee at Beth Israel Deaconess Medical Center (Boston, MA). All tissues were accompanied by pathology reports. Part of each sample was frozen in liquid nitrogen immediately after surgery and stored at ~80°C. The samples were made anonymous before the study. The International Union Against Cancer tumor-node-metastasis classification was used for pathology reports (11).

cDNA synthesis and microarray probe preparation. Total RNA was isolated using TRIzol reagent (Life Technologies, Rockville, MD) according to the manufacturer’s recommendations. Transcriptional profiling was done on HG-U133A Affymetrix GeneChips containing 22,283 genes. cDNA was prepared according to the manufacturer’s protocol. Total RNA (8 μg) was used in the first-strand cDNA synthesis with T7-(dT)24 primer (GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG-(dT)24) and SuperScript II (Life Technologies). The second-strand cDNA synthesis was carried out according to the manufacturer’s protocol.

Affymetrix GeneChip hybridization. cRNA (20 μg) was fragmented as described and hybridized with a preequilibrated HG-U133A Affymetrix chip at 45°C for 16 hours. After the hybridization cocktails were removed, the chips were washed, stained, and scanned according to the manufacturer’s protocols as described previously (12).

Microarray data preprocessing. Samples were analyzed using dChip (13), where a smoothing spline normalization method was applied before obtaining model-based gene expression values. The samples used for further analysis had 3/5 ratios (using glyceraldehyde-3-phosphate dehydrogenase and β-actin probes), present call percentage, or array outlier call percentage within 2 SDs (see Supplementary Data at http://www.bidmcgenomics.org/KidneyCancer/index.html). The 3/5 ratio gives an indication of the integrity of starting RNA, efficiency of transcription of cRNA, and/or in vitro transcription of cRNA. Present call percentage is the proportion of the mRNA transcripts that were expressed (“present”) in the sample. The percentage of array outlier probes was calculated with the dChip algorithm as described previously (13).

Unsupervised analysis. An average linkage hierarchical clustering technique (14) was used where the metric of similarity was Pearson’s correlation between the transcription profiles of the samples. This represented the degree of closeness between samples based on the total gene expression. We also investigated the projection of samples in three-dimensional using principal components analysis as a dimension-reduction method to explain the variation in the original data set with a reasonable number of variables (15).

Supervised analysis. We used supervised analysis of gene expression for finding differentially expressed genes, prediction, and pattern discovery. To find genes that are significantly differentially expressed between two groups of samples, we used combination of two criteria in dChip. If the 90% lower confidence bound (LCB) of the fold change between the two groups was >2 and the t test showed a significant difference (P < 0.001), the corresponding gene was considered to be differentially expressed. Tumor type specificity of gene expression was assessed using leave-one-out cross-validation with the weighted voting algorithm and significance of prediction accuracy was assessed by Fisher’s test on the confusion matrix and permutation tests (5). For gene pattern discovery, we employed a pattern recognition algorithm implemented by IBM’s Genes@Work software (16). This algorithm is a multivariate gene selection tool where a “pattern” is a subset of genes that have similar expression values in a subset of the samples in the phenotype set with respect to the control set. We identified gene expression patterns for clear cell RCC metastasis (P < 10^-15) that distinguish the metastasis samples from the primary tumor samples. Gene lists for secreted proteins and receptors were derived using our online gene annotation tool (http://www.bidmcgenomics.org) that queries multiple public databases with >90 identifiers, including cell location, biological function, protein families, etc. Additional support was derived from literature search where necessary. A more detailed discussion of the methods along with variables used can be found in Supplementary Data.

Results

Data analysis was carried on 59 primary tumors (22 clear cell RCC, 11 papillary RCC, 6 chromophobe RCC, 8 TCC, and 12 oncocytoma), 23 normal kidney tissue samples, and 1 nodal and 9 distant metastatic clear cell RCC samples. The clinical and pathologic characteristics of the samples are presented in Tables 1 and 2.

Transcriptional profiling of different histologies. We used the complete array expression values to build a dendrogram (Fig. 1), reflecting global expression similarities across the samples. Hierarchical clustering of all samples showed a clear distinction between cancerous and normal samples. Oncocytomas and chromophobe RCC grouped together in a subcluster with normal kidney samples. All other malignant tumors clustered in a separate branch in concordance with the known histologic type. Whereas all clear cell RCC samples grouped together, papillary RCC formed a separate subcluster with TCC.

We derived tumor type-specific profiles by comparing each of the five renal tumor types to the ensemble of remaining tumor types. We built 30-gene predictors for each such comparison and assessed their accuracy using leave-one-out cross-validation (5). At each iteration, a sample was left out and a 30-gene predictor was generated using the remaining samples to predict the left-out sample with the weighted voting method (5). All five subtype-specific signatures yielded 100% prediction accuracy. In Fig. 2 (and

| Table 1. Clinical and pathologic characteristics of all analyzed samples |
|----------------|----------------|
| **Histology** | **No. samples** |
| Clear cell RCC | 32 |
| Papillary RCC | 11 |
| Chromophobe RCC | 6 |
| TCC | 8 |
| Oncocytoma | 12 |
| Normal tissue | 23 |

NOTE: Representation of various renal tumor types.
Supplementary Data), we show 30-gene signatures created for each tumor type using all samples (5), which were further used for external validation (see below).

Transcriptional profiling of low-stage clear cell renal cell carcinoma tumor samples. Using dChip, we compared all (eight) clear cell T1 samples with the 23 normal kidney samples and obtained a list of 1,359 genes that are significantly up-regulated in T1 and 493 genes that are significantly down-regulated in T1.

Among the differentially expressed genes between normal kidney and T1 clear cell RCC, we identified a subset of up-regulated genes coding for secreted proteins (Fig. 3A) and cell surface receptors (Fig. 3B), two very important classes of targets for drug discovery and diagnostics. Many of these genes coding for secreted proteins are involved in various aspects of angiogenesis (e.g., vascular endothelial growth factor, angiopoietin-like 4, angiopoietin-like 2, basic fibroblast growth factor, placental growth factor, endothelial cell growth factor 1, endothelin 1, transforming growth factor-β1, transforming growth factor-β2, adrenomedullin, and ceruloplasmin). Various migration, invasion, and adhesion-related genes coding for secretory proteins (such as parathyroid hormone-like hormone, matrix metalloproteinase 14, and gonadotropin releasing hormone 2), chemokines (such as CCL5, CCL20, and CCL18), chemokine receptors (CX3CR1, CXCR4, and CXCR6), or tyrosine kinase receptors (e.g., FLT1, AXL, and epidermal growth factor receptor) were also up-regulated in the early-stage tumors (see Supplementary Data for complete lists).

Analysis of clear cell renal cell carcinoma progression and metastasis. To identify genes whose expression is progressively dysregulated during stepwise progression from normal kidney tissue (N) through early tumor stage (T1) to distant metastasis (M), we compared the corresponding clear cell RCC samples using dChip: N versus T1 and T1 versus M (Fig. 4). We identified 31 genes that are significantly differentially expressed in the same direction in both comparisons, thus showing a continuous deregulation pattern as the disease progresses. In Fig. 4, we present the colorgram for the 31 genes: 18 genes that are up-regulated in T1 compared with N and further up-regulated in M compared with T1 and 13 genes that are down-regulated in T1 compared with N and further down-regulated in M compared with T1. Examples of genes continuously deregulated on progression include caveolin 1, annexin A4, and lysyl oxidase (see Supplementary Data). Progressively dysregulated genes may play a role in driving tumors toward increased malignancy. Although expression of these genes may be increased or decreased in metastases, they are not metastasis specific.

The use of transcriptional profiling to build RCC-specific outcome signatures has been proposed previously (9). Our next goal was to find a clear cell RCC metastasis-specific gene pattern by comparing all (eight) clear cell T1 samples that have until now not developed any metastases (median, 32 months; range, 28-40 months) with all (nine) samples collected from distant organ metastases. Using Genes@Work, we identified a 155-gene pattern that predicted metastasis in this training set with 100% accuracy (Fisher’s P = 0.00004, permutation P < 0.001).

Our next goal was to determine whether this pattern preexists in more advanced primary clear cell RCC tumors that develop metastases. For this purpose, we tested the 155-gene metastasis signature on the nine primary clear cell RCC tumors, which had or later developed visceral metastases. Of these nine samples, which were not used in developing the signature, five tumors were predicted as “T1,” and four as “metastatic.” Four of the five tumors predicted as T1 did not have organ metastases at the time of surgery. All four tumors classified as metastatic had organ metastases at the time of surgery. Thus, this gene signature predicted the presence of organ metastases at the time of surgery with 88.9% accuracy in this independent group of tumors (see Supplementary Data), indicating differences in the biology between RCC tumors that present with distant metastases at the time of surgery and RCC tumors that only later develop metastases. We also did principal components analysis where samples classified as T1 and metastatic could be divided into two distinct areas within the three-dimensional space (see Supplementary Data). This was overall in concordance with our classification results as only the samples in the independent data set that had metastasis at the time of surgery fell in the same space as the metastatic samples.

External validation of gene signatures. To validate our gene signatures and gene signatures generated by other investigators, we tested several independent external data sets across different microarray platforms in four facets of analysis: validating our tumor type-specific gene signatures, comparing our clear cell RCC–associated gene lists with the lists available in the literature, testing our metastasis gene signature, and application of an external global metastatic signature to our samples.

Validation of tumor type–specific gene signatures. Higgins et al. (17) hybridized 23 clear cell RCC, 4 papillary RCC, 3 chromophobe RCC, 2 oncocytoma, 1 angiomylipoma, 5 granular cell carcinomas, and 3 normal samples to a 22,648-element cDNA microarray representing 17,083 different UniGene clusters. Because of the differences in gene content across the two microarray platforms, we were able to match 64 of our 120 tumor type–specific genes for clear cell RCC, papillary RCC, oncocytoma, and chromophobe RCC, 17, 14,
15, and 18 genes, respectively, on the Higgins et al. data set using Ensembl (version Hs 25.34e) and UniGene (Build Hs 177) databases (see Supplementary Data). Based on these 64 matching genes, we used dChip to hierarchically cluster the samples in the Higgins et al. data set (Fig. 5A). The resulting tree validated our gene signatures separating tumor types successfully where the clusters with clear cell RCC, papillary RCC, chromophobe RCC, granular cell carcinoma, and normal samples all show significant associations (P < 0.05) as calculated by dChip (see Supplementary Data). Similar to the results obtained by our data set, clear cell RCC samples were separated from normal, oncocyotma, chromophobe RCC, and papillary RCC. Oncocytomas grouped together with chromophobe RCC and were closer to normal than papillary RCC or clear cell RCC. We then split the Higgins et al. data set into two groups of 22, each containing 13 clear cell RCC samples. Using the 17 genes that matched from our clear cell RCC gene signature, we built a model on the first group of 22 (training set) and applied it on the second group of 22 (test set). The prediction accuracy on the test set was 91% (20 of 22 predicted correctly) with a Fisher’s P of 2.4 x 10^-8 (see Supplementary Data). These data most vividly show that our tumor type–specific gene signatures hold up in an independent external validation set across two completely different microarray platforms.

**Clear cell renal cell carcinoma–associated gene list.** Lenburg et al. compared 8 normal and 9 clear cell RCC samples using an unpaired t test (3-fold change and P < 0.03), and on the HG-U133A chip, they found 357 probes up-regulated in clear cell RCC versus normal and 669 probes down-regulated in clear cell RCC versus normal. When compared with our T1 clear cell RCC–associated gene lists of 1,359 up-regulated and 493 down-regulated genes, the intersection was 187 and 107, respectively, both of which showed a significant overlap (P < 10^-10), where the probability of observing such an overlap by chance is calculated using hypergeometric distribution (18, 19). In Fig. 5B, we show the overlap between the two studies and the intersecting gene lists are shown in Supplementary Data. These data show that there is a significant overlap between the gene lists generated in two independent microarray analyses, supporting the relevance of this gene set for clear cell RCC; however, there are also significant differences that can be due to a combination of several factors, including the difference in algorithms used for deriving the signal values (we used the dChip algorithm and Lenburg et al. applied the MAS5 algorithm) and the difference in sample composition (we used only T1 stage clear cell RCC versus normal kidney to obtain differentially expressed genes and Lenburg et al. used clear cell RCC samples with undefined tumor stages).

**Validation of the metastasis gene signature.** Sultmann et al. hybridized 19 primary RCC samples that had metastasis at the time of diagnosis (M1) and 17 that did not (M0) on custom cDNA microarrays containing 4,207 genes and expressed sequence tags (20). We matched 41 of our 155 metastasis-associated genes across the two different microarray platforms.
in the Sultmann et al. data set using Ensembl (version Hs 25.34e) and UniGene (Build Hs 177; see Supplementary Data). Using dChip, we clustered these 36 samples (Fig. 5C) based on the 41 matched genes. The two groups, M₀ and M₁, were separated into two significantly distinct clusters (P < 0.05 for each branch as calculated by dChip), independently validating our notion about the presence of the metastasis signature in primary clear cell RCC tumors with metastasis at the time of diagnosis but not in tumors that did not have metastasis at the time of surgery. Again, as expected due to the difference in microarray platforms and the resulting small overlap of genes across the two platforms (41 of 155), separation of the two phenotypes was not perfect, but despite these limitations, our metastasis gene signature resulted with statistical significance in this independent data set.

External global metastasis gene signature. Using Affymetrix HG-U95A arrays, Ramaswamy et al. compared 12 metastatic and 64 primary adenocarcinomas from several tumors (lung, breast, prostate, colorectal, uterus, and ovary) to find a 128-gene metastatic signature that correlates with poor clinical outcome (21). Because this global metastatic signature has not been evaluated on RCC, we decided to determine its applicability for RCC.

We used 169 probes (corresponding to those 128 predictor genes) from the HG-U95A array (that was used in the previous work) and mapped them to the HG-U133A array using Affymetrix “best matched” criteria for same-species comparison guideline. The resulting 123 analogues were used to build an unsupervised hierarchical tree with our T₁ and distant metastatic clear cell RCC samples as well as our >T₁ samples with or without metastases at the time of surgery. The T₁ samples have until now not developed metastases. The resulting tree showed that the majority of distant metastases and the tumors with metastasis at the time of surgery correctly clustered in their corresponding known classes in a separate branch from the majority of T₁ samples and the T₂ and T₃ RCC samples that only later developed metastases (see Fig. 5D). These findings suggest that the previously reported (21) global metastatic gene signature can successfully distinguish distant metastasis from tumors that do not develop metastases in RCC and can further distinguish between RCC with metastasis at the time of surgery and tumors that later developed metastases. Due to the lack of sufficient follow-up, no conclusion about clinical outcome can be drawn at the moment. These data also confirm the results obtained with our metastasis signature that separated RCC with metastasis at the time of diagnosis from tumors that later developed metastases and further support our hypothesis that these tumors are biologically distinct.

Discussion

Despite great refinements in surgical technique, perioperative care, and increased knowledge of immunobiology, there are major limitations in managing RCC. This becomes particularly obvious when trying to identify prognostic markers of outcome or when treating patients with metastatic disease. Currently available clinical prognostic markers (such as stage or performance status) are useful, but substantial outcome variability exists even within patients stratified for these prognostic factors. Molecular markers of RCC have also been proposed (22), but it is questionable whether they
adequately address the heterogeneity of this disease when studied in isolation. Current knowledge of RCC mostly derives from the analysis of the most common subtype (clear cell RCC) and fails to address ∼30% of the remaining subtypes, as the sample size of the less common subtypes is often the limiting factor for statistically robust statements in most studies. Furthermore, even within a subtype, there can be more and less aggressive molecular variants leading to different clinical outcomes as proposed recently for clear cell RCC (9). Our aim was to detect these molecular differences in various RCC subtypes and during tumor progression. We expect that our findings will be useful in individualizing diagnosis and therapy in the future and in deciphering underlying biological differences.

In our study, we analyzed global gene expression of the three most common RCC subtypes (clear cell, papillary, and chromophobe RCC), TCC, and oncocytoma in comparison with normal kidney tissue. TCC and oncocytoma are important in the differential diagnosis of RCC and, at initial radiological presentation, often indistinguishable from it. Whereas TCC is a highly malignant tumor arising from the urothelium of the renal pelvis, oncocytomas are benign tumors originating from the intercalating cells of the collecting duct (23). Hierarchical clustering showed a perfect distinction of RCC, TCC, oncocytoma, and normal kidney tissue. Whereas the clear cell and papillary RCC subtypes clustered together with TCC, chromophobe RCC grouped with normal kidney tissue and oncocytoma. This finding is in concordance with the more benign nature of chromophobe RCC, which has been shown recently to have the highest 5-year disease-specific survival among all RCC subtypes (24). Furthermore, it potentially reflects the common site of origin (collecting duct) for oncocytoma and chromophobe RCC (25). In contrast, clear cell RCC and papillary RCC did not group together despite their common origin in the proximal tubule of the renal cortex. Their reported difference in aggressiveness and outcome may be reflected in this finding (24, 26). Given that RCC and TCC are distinct tumor entities, the clustering of papillary RCC and TCC was surprising and unexpected. Examination of a previous microarray publication with a smaller number of papillary RCC and TCC reveals a similar clustering of papillary RCC with TCC but not with clear cell RCC (10). This interesting finding needs to be further confirmed in larger, more specifically designed studies.

Our novel analysis extends previous microarray studies addressing the different histologic subtypes of RCC. These studies reported genes differentially expressed between different subtypes but came short of reporting the accuracy of their specific gene signatures (10, 17). In our study, we provide subtype-specific signatures with 100% accuracy. We also validated our subtype gene signatures across two different microarray platforms on an independent external data set by Higgins et al. (17). This is one of the few times that this cross-platform validation has been done. Nevertheless, due to the difference between the microarray platforms used, only a subset of the genes overlapped between the two data sets. Despite this limitation of gene overlaps and the difference in the microarray platform itself, our gene signature did extremely well and separated all tumor samples of the external data accurately, further validating our analysis and demonstrating the cross-platform and cross-laboratory performance. Although the tumor specificity of our gene lists might serve for differential diagnoses in selected cases, we believe these sets of genes mainly have the potential of elucidating unique biological pathways inherent to the different renal tumors that can be explored for discovery of novel molecular drug targets and prognostic biomarkers.
The focus of our further analysis was directed toward molecular characterization of clear cell RCC tumorigenesis and progression. We first analyzed the transcriptional profiles with respect to early and late phenotypic stages of clear cell RCC and subsequently defined a characteristic set of genes for the T1 stage. Among the differentially expressed genes in T1 clear cell RCC tumors, we identified a subset of genes coding for secreted proteins and cell membrane receptors that may be used in the future as biomarkers or molecular targets for drug development. Many of these genes coding for secreted proteins are involved in various aspects of angiogenesis, such as vascular endothelial growth factor, angiotatin-like 2, angiotatin-like 4, basic fibroblast growth factor, placental growth factor, endothelial cell growth factor 1, transforming growth factor-β1, transforming growth factor-β2, adrenomedullin, and ceruloplasmin, and have been associated with RCC (27–32). This may reflect the high vascularity of clear cell RCC, possibly due to at least partially mutated von Hippel-Lindau, and is in contrast to the other subtypes of RCC, oncocytoma, and TCC that only show sporadic up-regulation of angiogenic factors. Among up-regulated receptors, tyrosine kinase receptors (such as FLT1 or epidermal growth factor receptor), chemokine receptors (such as CXCR4), and apparent autocrine loops (such as transforming growth factor-β1/epidermal growth factor receptor) are particularly interesting, because they all have been implicated in various aspects of development or progression of RCC or other types of cancer (27, 33–38). Thus, the angiogenic phenotype seems to be already expressed at early stages of clear cell RCC and is likely to play an important role in tumor development and progression. Further elucidation of the precise identity and role of the various genes in these networks may allow development of new therapeutic models for various clear cell RCC tumor stages. Our study represents a first step toward this direction by providing evidence for a concerted up-regulation of several angiogenesis and growth factor pathway genes in clear cell RCC, suggesting the potential therapeutic utility of angiogenesis inhibitors, tyrosine kinase inhibitors, and growth factor or receptor antagonists. In addition, novel approaches with small-molecule antagonists of a chemokine receptor, such as CXCR4, which have been successfully applied in animal models for other tumors, carry the potential of being effective in RCC (39).

We further identified within clear cell RCC a subset of genes progressively dysregulated from normal kidney tissue through T1 stage to metastasis. These progressively dysregulated genes do not represent metastasis-specific genes, because they are already up-regulated or down-regulated at the T1 stage compared with normal kidney but are of significant biological interest with regard to tumor progression and may be involved in driving the progression toward increased malignancy. Some of these genes (e.g., caveolin 1, hsp 72, and annexin A4) have already been associated with RCC and aggressive tumor behavior (40–42). Genes continuously dysregulated throughout RCC progression might be interesting diagnostically and therapeutically alike.

Metastatic RCC has a particularly poor prognosis with only ~9% 5-year survival and available treatment options are associated with response rates of 15% to 20% with a median response duration of 54 months (43). Our goal was to identify a gene pattern that best distinguishes metastatic samples and use it to explore the metastatic potential of primary tumors. This novel signature distinguished primary tumors very accurately in two groups based on the presence of metastases at the time of surgery, in two independent validation sets, one of which from another study done on a different microarray platform. This issue had not been addressed previously by other important reports focusing on advanced RCC (9, 44). Additionally, we evaluated, to our knowledge for the first time in RCC, a previously reported global metastatic signature that had been derived from various non-RCC solid tumors (21). An unsupervised classification of our clear cell RCC samples using this global gene signature correctly separated the majority of distant metastases and RCC tumors that only later developed metastasis signature classifies most of the primary tumor samples with metastasis at the time of surgery together with distant metastases and RCC tumors that only later developed metastasis together with the T1 tumors, indicating that there is
an intrinsic biological difference between tumors that have developed distant metastasis at the time of surgery and tumors that apparently are less aggressive and do not present with metastases at the time of surgery. These findings suggest that there is an inherent gene signature in the primary tumor that may be associated with a highly aggressive course. Moreover, it is in partial agreement with the hypothesis generated by Ramaswamy et al., suggesting that the metastatic potential of human tumors is encoded in the bulk of the primary tumor (21). However, at least in RCC, our metastatic signature and the global metastatic signature by Ramaswamy et al. suggest that a subset of more aggressive RCC tumors (metastasis at the time of surgery) indeed encodes the metastatic potential in the bulk of the primary tumor, whereas another subset of less aggressive tumors (metastasis developed only later) does not encode the metastatic signature in the primary tumor. From a clinical perspective, the ability to identify patients at high risk for distant metastasis may lead to more extensive imaging evaluation of these patients at diagnosis and subsequently to initiation of further therapy. According to some series, a third of patients with bone metastasis did not have musculoskeletal pain at initial diagnosis and would have been missed, whereas ~50% would have been missed because of normal levels of serum alkaline phosphatase had the indication for a bone scintigraphy relied on symptoms or elevated serum alkaline phosphatase, respectively, as often suggested (45–47). In addition, potential intrinsic biological differences between more and less aggressive RCC tumors may suggest different therapeutic approaches with different molecular targets.

The feasibility of applying microarray data generated on a different platform and by a different research group has the potential to further facilitate translational research by directly implementing existing microarray results into new studies. We have used four different external data sets using various different microarray platforms to either validate our data or validate external data and we have shown that, despite all the limitations, validation on independent external data sets is feasible.

Our clear cell RCC metastasis gene signature contains several genes that have been associated with cancer progression, metastasis, or poor outcome in RCC or other cancers (matrix metalloproteinase 14, topoisomerase IIα, glucosyl ceramide synthase, and ADAM 20; refs. 48–51). Particularly interesting is the up-regulation of topoisomerase IIα in conjunction with glucosylceramide synthase. Topoisomerase IIα has been known to be expressed in higher-grade RCC (49). Although topoisomerase IIα inhibitors (such as doxorubicin and etoposide) are effective in a variety of tumors, their efficacy in RCC is very limited (52–54). Interestingly, the glucosylceramide synthase enzyme,

![Color map of the 31 genes that are significantly differentially expressed in N versus T, and T versus M. Expression profiles of 18 genes are positively correlated with disease progression, whereas expression profiles of 13 genes are negatively correlated with disease progression. Genes are represented in rows; samples are represented in columns where the title for a sample group indicates the beginning of that particular group of samples. Red and green indicate high and low expression, respectively.](image-url)
which has been implicated in tumor cell resistance to the topoisomerase inhibitor doxorubicin, is also among the up-regulated genes suggesting a mechanism how clear cell RCC tumors could escape cell death in response to topoisomerase inhibitors. OGT2378, an imino sugar inhibitor of glucosylceramide synthesis, was shown recently to inhibit tumor growth of melanoma cells (55). Blocking glucosylceramide synthase may also overcome resistance to topoisomerase inhibitors and pave the way to effective combination therapy. Other less well known genes within the metastatic signature, which have not been associated with RCC in the past, such as the transcription factor POU6F1, may deserve evaluation in future studies (56). We have used several pathway analysis software tools to integrate the metastatic genes into biological pathways, revealing a complex interaction between many of the metastatic genes and several nodal points where multiple metastatic genes act on (see Supplementary Data). In particular, the three tumor suppressor and oncogene pathways involving p53, c-myc, and c-fos as nodal points seem to be targeted by various metastatic gene products. Multiple signal transduction pathways, such as the mitogen-activated protein kinase, nuclear factor-κB, and phosphatidylinositol 3-kinase/AKT pathways, cell survival, motility, and proliferation seem to be modulated by metastasis genes.

Historically, RCC was considered a single entity with various histologic appearances. Today, RCC is more accurately recognized as a family of cancers resulting from distinct genetic abnormalities with unique morphologic features but a common derivation from the renal tubular epithelium (57). We and others (8–10) have shown that transcriptional profiling is a powerful tool for better appreciation of the subtle phenotypic differences among various renal tumors in general and RCC in particular. Data generated in these studies, if prospectively validated, should aid in the development of improved treatment approaches for RCC.

Acknowledgments

We thank Vikas Sukhatme and Ananth Karumanchi for comments on the article and Christa Blumenberg for kind help and assistance.