# Gene Expression Profiling Separates Chromophobe Renal Cell Carcinoma from Oncocytoma and Identifies Vesicular Transport and Cell Junction Proteins as Differentially Expressed Genes

Stephen Rohan,[1] Jiangling J. Tu,[1] Jean Kao,[1] Piali Mukherjee,[2] Fabien Campagne,[2] Xi K. Zhou,[3] Elizabeth Hyjek,[1] Miguel A. Alonso,[4] and Yao-Tseng Chen[1]

**Abstract**

**Purpose:** To compare gene expression profiles of chromophobe renal cell carcinoma (RCC) and benign oncocytoma, aiming at identifying differentially expressed genes.

**Experimental Design:** Nine cases each of chromophobe RCC and oncocytoma were analyzed by oligonucleotide microarray. Candidate genes that showed consistent differential expression were validated by reverse transcription-PCR using 25 fresh-frozen and 15 formalin-fixed, paraffin-embedded tumor samples. Immunohistochemical analysis was also done for two selected gene products, claudin 8 and MAL2.

**Results:** Unsupervised hierarchical clustering separated the chromophobe RCC and oncocytoma into two distinct groups. By a combination of data analysis approaches, we identified 11 candidate genes showing consistent differential expression between chromophobe RCC and oncocytoma. Five of these genes, *AP1M2*, *MAL2, PROM2, PRSS8*, and *FLJ20171*, were shown to effectively separate these two tumor groups by quantitative reverse transcription-PCR using fresh tissue samples, with similar trends seen on formalin-fixed tissues. Immunohistochemical analysis revealed selective expression of MAL2 and claudin 8 in distal renal tubules, with MAL2 antibody showing differential expression between chromophobe RCC and oncocytoma. Functional analyses suggest that genes encoding tight junction proteins and vesicular membrane trafficking proteins, normally expressed in distal nephrons, are retained in chromophobe RCC and lost or consistently down-regulated in oncocytoma, indicating that these two tumor types, believed to be both derived from distal tubules, are likely distinctive in their histogenesis.

**Conclusions:** We showed that chromophobe RCC and oncocytoma are distinguishable by mRNA expression profiles and a panel of gene products potentially useful as diagnostic markers were identified.

Renal cell carcinoma (RCC) is a heterogeneous group of malignancy, and clear cell, papillary, and chromophobe RCC are the major subtypes (1). Of these, the chromophobe RCC, constituting 5% to 10%, is the least common and has morphologic features that often overlap with oncocytoma, a benign neoplasm. The distinction between these two tumors is clinically important, as chromophobe RCC, although considered to have better prognosis than conventional clear cell carcinoma (2), is malignant and can potentially be aggressive.

The similarity between chromophobe carcinoma and oncocytoma likely reflects their shared histogenesis from the intercalated cells of the distal tubules, a notion postulated based on ultrastructural findings (3). This similarity was further supported by the recent cDNA or oligonucleotide microarrays in which these two tumor types could not be reliably separated. In contrast, chromophobe carcinoma and oncocytoma as a group showed distinctive microarray profiles, easily separated from clear cell and papillary RCC (4–8).

Despite these similarities, biological differences between these two entities are unequivocal. Cytogenetic evidence was most compelling, with chromophobe RCC showing hypodiploidy, often with monosomy of chromosomes 1, 2, 6, 10, 13, 17, and 21 (9–11). In comparison, oncocytomas often display mixed karyotypes, with loss of chromosomes Y and 1 often observed (12).

Additional differences between chromophobe RCC and oncytoma were observed in immunohistochemical studies. For example, kidney-specific cadherin and epithelial cell adhesion molecule were shown in one study to be expressed in all, or almost all, chromophobe RCC, but rarely in oncocytoma (13). Although this finding was challenged by later studies (14), other markers such as cytokeratin 7, parvalbumin, and claudin 7 were

**Authors' Affiliations:** [1]Department of Pathology and Laboratory Medicine, [2]Institute for Computational Biomedicine, and [3]Department of Public Health, Weill Medical College of Cornell University, New York, New York; and [4]Centro de Biología Molecular "Severo Ochoa," Consejo Superior de Investigaciones Científicas and Universidad Autónoma de Madrid, Madrid, Spain

Received 5/24/06; revised 8/8/06; accepted 8/31/06.

**Grant support:** Ministerio de Educación y Ciencia grants BMC2003-03297, BFU2006-01925, and GEN2003-20662-C07-02 (M.A. Alonso) and an institutional grant from the Fundación Ramón Areces to Centro de Biología Molecular "Severo Ochoa."

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

**Requests for reprints:** Yao-Tseng Chen, Department of Pathology, Weill Medical College of Cornell University, Box 69, 1300 York Avenue, New York, NY 10021. Phone: 212-746-6472; Fax: 212-746-8166; E-mail: ytchen@med.cornell.edu.



doi:10.1158/1078-0432.CCR-06-1268

**Table 1.** Differentially expressed genes (67 probe sets, corresponding to 57 genes) with *P* values of <0.05 and fold change >1.5

| Common name | GenBank ID | Fold change | *P* | Description |
|---|---|---|---|---|
| **Genes overexpressed in chromophobe RCC** | | | | |
| ***TMC4*** | **BE645551** | **5.902** | **0.000116** | **transmembrane channel-like 4** |
| ***AP1M2*** | **NM_005498** | **5.196** | **0.0119** | **adaptor-related protein complex 1, μ2 subunit** |
| ***AP1M2*** | **AA910946** | **3.560** | **0.0119** | **adaptor-related protein complex 1, μ2 subunit** |
| *EPB41L4B* | NM_019114 | 2.862 | 0.0119 | erythrocyte membrane protein band 4.1 like 4B |
| *CENTA1* | AW050627 | 2.428 | 0.0119 | centaurin, alpha 1 |
| (N.A.) | AW302207 | 2.020 | 0.0119 | transcribed sequences |
| *SH3MD2* | AI686957 | 1.636 | 0.0119 | SH3 multiple domains 2 |
| ***FLJ20171*** | **NM_017697** | **5.786** | **0.0127** | **hypothetical protein FLJ20171** |
| ***CLDN8*** | **AL049977** | **11.890** | **0.0129** | **claudin 8** |
| *SPINT1* | NM_003710 | 5.799 | 0.0229 | serine protease inhibitor, Kunitz type 1 |
| *C14orf114* | NM_018199 | 1.581 | 0.0229 | chromosome 14 open reading frame 114 |
| ***MAL2*** | **AL117612** | **29.340** | **0.0238** | **mal, T-cell differentiation protein 2** |
| *MANBA* | NM_005908 | 1.929 | 0.0238 | mannosidase, βA, lysosomal |
| *SLC27A1* | BF056007 | 1.859 | 0.0238 | solute carrier family 27 (fatty acid transporter), member 1 |
| *C14orf87* | AA133341 | 1.501 | 0.0238 | chromosome 14 open reading frame 87 |
| (N.A.) | AI191905 | 4.740 | 0.024 | transcribed sequences |
| *LOC196264* | AA772172 | 1.809 | 0.024 | hypothetical protein LOC196264 |
| *MGC21874* | AI862537 | 1.754 | 0.024 | transcriptional adaptor 2 (ADA2 homologue, yeast)-β |
| ***TJP3*** | **NM_014428** | **1.573** | **0.024** | **tight junction protein 3 (zona occludens 3)** |
| ***CLDN7*** | **NM_001307** | **5.559** | **0.0298** | **GABA(A) receptor-associated protein** |
| *CDS1* | NM_001263 | 4.756 | 0.0298 | CDP-diacylglycerol synthase 1 |
| ***TJP3*** | **AC005954** | **2.893** | **0.0298** | **tight junction protein 3 (zona occludens 3)** |
| *DKFZP566J2046* | AW070436 | 2.010 | 0.0298 | hypothetical protein DKFZp566J2046 |
| *EPS8L1* | BC004907 | 1.650 | 0.0298 | EPS8-like 1 |
| ***FLJ21918*** | **NM_024939** | **1.945** | **0.0333** | **hypothetical protein FLJ21918** |
| *LRRC1* | NM_018214 | 2.236 | 0.0379 | leucine-rich repeat containing 1 |
| *STX3A* | BE966922 | 2.488 | 0.0383 | NIH_MGC_72 *Homo sapiens* cDNA clone IMAGE:3915610 |
| ***FLJ20171*** | **BF001941** | **19.180** | **0.0405** | **hypothetical protein FLJ20171** |
| *PVALB* | NM_002854 | 4.815 | 0.0405 | parvalbumin |
| (N.A.) | AI038402 | 2.501 | 0.0438 | Transcribed seq. similar to A57377 transcription factor NFATx |
| ***CAPN1*** | **NM_005186** | **1.807** | **0.044** | **calpain 1, (μ/I) large subunit** |
| *C14orf108* | NM_018229 | 2.445 | 0.0442 | chromosome 14 open reading frame 108 |
| *CDS1* | AW304313 | 2.334 | 0.046 | CDP-diacylglycerol synthase 1 |
| *TACSTD1* | NM_002354 | 3.307 | 0.0461 | tumor-associated calcium signal transducer 1 |
| *CA2* | M36532 | 2.733 | 0.0461 | carbonic anhydrase II |
| *FLJ36445* | AA827649 | 2.099 | 0.0461 | hypothetical protein FLJ36445 |
| *SLC16A7* | AW975728 | 2.238 | 0.047 | solute carrier family 16, member 7 |
| *MDA5* | NM_022168 | 1.671 | 0.047 | melanoma differentiation associated protein-5 |
| *C14orf108* | AW137526 | 1.984 | 0.0483 | chromosome 14 open reading frame 108 |
| *C14orf125* | BF435286 | 1.682 | 0.0485 | chromosome 14 open reading frame 125 |
| *SH3YL1* | NM_015677 | 2.916 | 0.0493 | SH3 domain containing, Ysc84-like 1 |
| ***HOOK2*** | **NM_013312** | **2.629** | **0.0493** | **hook homologue 2 (*Drosophila*)** |
| *FLJ34633* | AA573775 | 2.072 | 0.0494 | hypothetical protein FLJ34633 |

(*Continued on the following page*)

also described as preferentially expressed in chromophobe RCC over oncocytoma (4, 15–17). In contrast, S100 protein was found to be preferentially expressed in oncocytoma (18). The clinical diagnostic usefulness of these markers, however, needs to be further confirmed by additional studies.

Based on these observations, it is believed that these two tumors are biologically distinct, but no unequivocal distinguishing markers have been defined. In this study, we did oligonucleotide microarray analysis to search for such biological markers, and the diagnostic potential of these new markers was explored.

## Materials and Methods

***Tissue specimens.*** Tissue specimens were obtained from Department of Pathology at the Weill Medical College of Cornell University following an Institutional Review Board–approved protocol. The H&E slides on all cases were reviewed by one of us (J.J.T.), and only histologically unequivocal cases were included.

***RNA extraction from fresh-frozen and paraffin-embedded tissues.*** Total RNA was extracted from fresh tissues using RNeasy mini kit (Qiagen, Valencia, CA). Approximately 30 mg of fresh tissues were used for each sample. For extraction of RNA from paraffin-embedded tissue, the Optimum FFPE RNA isolation kit (Ambion, Austin, TX) was used with materials derived from four 8-μm sections. The nontumor areas on the slides were manually removed with surgical blades and the remaining tissue on the slide was scraped into an Eppendorf tube for RNA extraction.

***Microarray experiments.*** RNA was reverse transcribed and *in vitro* transcribed and biotin labeled using Affymetrix one-cycle cDNA synthesis and IVT labeling kits (Affymetrix, Santa Clara, CA). Following biotinylation and fragmentation, hybridization was done against the Affymetrix Human HG-U133 plus 2.0 GeneChips according to the manufacturer's directions. The HG-U133 Plus 2.0

**Table 1.** Differentially expressed genes (67 probe sets, corresponding to 57 genes) with *P* values of <0.05 and fold change >1.5 (Cont'd)

| Common name | GenBank ID | Fold change | *P* | Description |
|---|---|---|---|---|
| **Genes overexpressed in oncocytoma** | | | | |
| ***APOE*** | **NM_000041** | **−3.623** | **0.0119** | **apolipoprotein E** |
| ***APOE*** | **AI358867** | **−2.278** | **0.0129** | **apolipoprotein E** |
| *BNIP3* | U15174 | −2.198 | 0.0229 | BCL2/adenovirus E1B 19 kDa-interacting protein 3 |
| *CUGBP2* | U69546 | −3.676 | 0.0229 | CUG triplet repeat, RNA binding protein 2 |
| ***DOCK1*** | **AA599017** | **−2.353** | **0.0238** | **dedicator of cytokinesis 1** |
| *CUGBP2* | N36839 | −3.175 | 0.0238 | CUG triplet repeat, RNA binding protein 2 |
| ***APOE*** | **N33009** | **−3.984** | **0.0238** | **apolipoprotein E** |
| *ABCC3* | NM_020037 | −1.524 | 0.024 | ATP-binding cassette, sub-family C (CFTR/MRP), member 3 |
| ***DOCK1*** | **NM_001380** | **−2.151** | **0.024** | **dedicator of cytokinesis 1** |
| *HLA-C* | M90685 | −1.600 | 0.0383 | HLA-G histocompatibility antigen, class I, G |
| ***GDI2*** | **D13988** | **−1.757** | **0.0383** | **GDP dissociation inhibitor 2** |
| *BNIP3* | NM_004052 | −2.066 | 0.0383 | BCL2/adenovirus E1B 19 kDa-interacting protein 3 |
| *FMNL3* | AW027431 | −1.597 | 0.0438 | formin-like 3 |
| *TAF15* | NM_003487 | −1.502 | 0.0439 | TAF15 RNA polymerase II, TBP-associated factor |
| *OK/SW-cl.56* | BC001002 | −2.024 | 0.046 | β5-tubulin |
| *FLJ21069* | NM_024692 | −1.508 | 0.0485 | hypothetical protein FLJ21069 |
| *HLA-G* | M90684 | −1.529 | 0.0485 | HLA-G histocompatibility antigen, class I, G |
| *HLA-B* | D83043 | −1.776 | 0.0485 | major histocompatibility complex, class I, B |
| *HLA-F* | AW514210 | −1.859 | 0.0485 | major histocompatibility complex, class I, F |
| *NBL1* | NM_005380 | −2.370 | 0.0485 | neuroblastoma, suppression of tumorigenicity 1 |
| *CPNE2* | AW170571 | −2.899 | 0.0485 | copine II |
| *MAPRE3* | BG222594 | −3.584 | 0.0485 | microtubule-associated protein, RP/EB family, member 3 |
| (N.A.) | M80469 | −1.912 | 0.0493 | heavy chain; Human MHC class I HLA-J gene |
| *RAD51C* | NM 002876 | −1.590 | 0.0494 | RAD51 homologue C (*S. cerevisiae*) |

NOTE: Genes chosen for validation (Table 3) and/or identified by pathway analysis (Table 4) are shown in bold.

GeneChips (Affymetrix) contain 54,675 probe sets that correspond to 38,500 genes (and >47,400 transcripts).

***Microarray data analysis.*** The microarray data were analyzed with GeneSpring 7.2 Software (Agilent Technologies, Santa Clara, CA). Raw image data were preprocessed using the RMA algorithm (19). Probe set data were median normalized per chip. Differential expression between oncocytoma and chromophobe RCC samples was assessed by ANOVA (Welch *t* test) with Benjamini and Hochberg multiple testing correction to control for the false discovery rate (20). Probe sets with ANOVA *P* values of <0.05 and fold changes of >1.5 were considered significant. The samples were then clustered using the GeneSpring hierarchical clustering algorithm.

***Supervised learning.*** Support vector machine (21) has been found to outperform other machine learning approaches (e.g., artificial neural networks; ref. 22). Signals of probe sets found to be differentially expressed were used as features of the support vector machine, and the histologic assignments of tissues were used as class label (i.e., chromophobe RCC class −1, oncocytoma class +1). The decision function of a support vector machine is a function of the feature vector and of variables learned during training (23). Among these variables, the vector associates one weight to each feature used to train the support vector machine and indicates how strongly the feature will govern the decision of the support vector machine.

Features were mean normalized across samples and the performance of the support vector machine was evaluated with the leave-one-out protocol. Thorsten Joachims' SVMlight program was used to perform support vector machine training (23) with a linear kernel and default values. After training, the 20 probe sets with the highest absolute support vector machine weight were considered to carry the information most useful for the prediction task. The significance of each classification was assessed using a label permutation test similar to that described by Mukherjee et al. (24), with 1,000 label permutations.

***Reverse transcription-PCR.*** Conventional and quantitative reverse transcription-PCR (RT-PCR) were done as described (25). Conventional RT-PCR was done with 30 amplification cycles to emulate a semi-quantitative assay, whereas quantitative RT-PCR assays were run for 45 cycles. Both 18S rRNA and glyceraldehyde-3-phosphate dehydrogenase were used as endogenous controls. All fresh-tissue RNA was found to be similar in RNA quality, whereas a wider variation was seen in formalin-fixed tissues. The 18S rRNA was found to be a more reliable control and all quantitative RT-PCR data were normalized against the 18S rRNA, with the mRNA expression value expressed as $\Delta Ct = Ct_{experimental\ gene} - Ct_{18S\ rRNA}$.

***Immunohistochemical analysis.*** Immunohistochemical analysis was done for claudin 8 and MAL2 protein. Anti–claudin 8 antibody (GeneTex, San Antonio, TX), a rabbit polyclonal antibody, was used at 1:200 dilution, with a Techmate$^{500}$ automated immunostainer (Ventana Medical Systems, Inc., Tuscan, AZ). The staining was done according to a modified MIP protocol using Envision+ horseradish peroxidase rabbit detection system (DakoCytomation, Carpinteria, CA). Anti-MAL2 antibody, mouse hybridoma 9D1 (26, 27), was used at 1:100 dilution, following previously published procedures (26, 27).

***Functional analysis.*** We searched Gene Ontology functional categories and Kyoto Encyclopedia of Genes and Genomes functional pathways for statistically enriched clusters/groups among the differentially expressed genes identified in this study. We used the EASE software (28) with human Unigene and GenBank identifiers. The *P* value reported is the EASE score (28). Differentially expressed genes were also analyzed using Ingenuity Pathways Analysis (Ingenuity Systems).[5]

## Results

***Identification of distinguishing genes.*** Several complementary data analysis approaches were used to identify genes

[5] http://www.ingenuity.com.

differentially expressed between the chromophobe RCC and oncocytoma tumor groups. Using normalized microarray data, ANOVA, and a multiple testing correction, we identified 67 probe sets (corresponding to 57 genes) with ANOVA *P* values of <0.05 and expression changes >1.5-fold (Table 1). Of these, 38 genes were overexpressed in chromophobe RCC, whereas only 19 were overexpressed in oncocytoma.

Unsupervised clustering using all genes in microarray separated the two groups when the data were not preprocessed with the RMA algorithm, but this distinction was lost when RMA algorithm was used (data not shown). However, even with RMA preprocessing, unsupervised clustering using the 67 differentially expressed probe sets accurately separated the two groups (Fig. 1). The two eosinophilic variants of chromophobe RCC did not cluster together.

Further evaluation, however, revealed significant difference in the intragroup expression consistency of the genes shown in Table 1. For example, MAL2 showed very tight and nonoverlapping ranges of expression between oncocytoma group (normalized microarray value, 0.304-1.601) and chromophobe RCC group (10.84-26.34), indicating MAL2 as a promising marker. In contrast, parvalbumin (PVALB), although showing a high fold-change value (4.815), had variable expression in oncocytoma (normalized value, 0.397-81.89) that overlapped with the chromophobe RCC cases (range, 55.19-135.1). These data indicate that PVALB would not be a reliable marker, and this was indeed shown in our previous RT-PCR study (25). By manually evaluating this intragroup variability and intergroup range differences, seven genes, *AP1M2*, *MAL2*, *FLJ20171*, *TMC4*, *CLDN7*, *CLDN8*, and *APOE*, emerged as the best candidate genes for distinguishing chromophobe RCC and oncocytoma. *APOE* showed higher expression in oncocytoma, with all other genes being higher in chromophobe RCC.

Because multiple testing corrections can sometimes be too conservative, we tested how this correction affected our identification of gene markers. To do so, we analyzed microarray data with ANOVA *P* values of <0.001 and fold change >2 without applying a multiple testing correction. This approach generated a gene list of 1,254 genes, and the top 80 genes were manually evaluated for their consistency in expression as above. This search confirmed *AP1M2*, *MAL2*, *CLDN7*, *CLDN8*, and *FLJ20171* as distinguishing markers, but not *TMC4* or *APOE*. In addition, it identified *CEL*, *KRT7*, *PRSS8*, and *PROM2* as candidate genes, all with higher expression in chromophobe RCC than in oncocytoma.

In addition to these univariate analyses, we evaluated a support vector machine classifier to see which combination of probe set expression levels best predict the tumor group (see Materials and Methods). The leave-one-out measures and the significant label permutation test *P* values indicate that a support vector machine can be trained to reliably predict the tumor histology. We then analyzed the trained support vector machine to identify the probe sets that carried the most weight in the trained decision function of the support vector machine. This analysis confirmed *APOE*, *CLDN8*, and *MAL2* in the top 20 genes and provided additional candidates (Table 2).

Combining results from these different data analysis approaches, 11 most promising genes were identified (Table 3). Of these, claudin 7 (*CLDN7*) was not further analyzed, as it was evaluated recently by Schuetz et al. (4) immunohistochemically with suboptimal results. Claudin 8 (*CLDN8*) was a single-exon

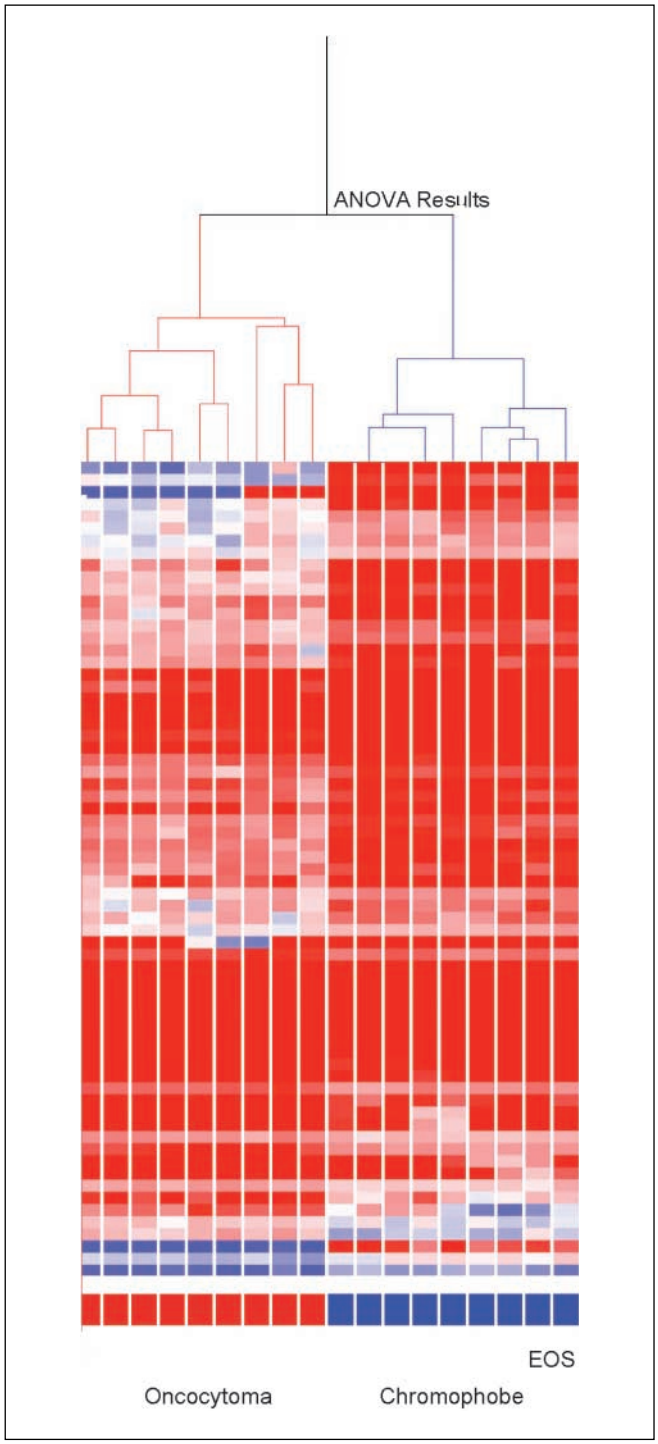


**Fig. 1.** Unsupervised clustering of the 67 differentially expressed probe sets separated chromophobe RCC from oncocytoma. Two eosinophilic variants of chromophobe RCC did not cluster together.

(intronless) gene (RefSeq. NM_199328) and was excluded for RT-PCR evaluation, as contaminating genomic DNA would result in the same PCR product, complicating quantitative RT-PCR evaluation of mRNA expression. Instead, claudin 8 was evaluated immunohistochemically (see below).

***Validation by RT-PCR on fresh tissues.*** The expression of the remaining nine genes was evaluated using RNA extracted from

**Table 2.** Genes identified by multivariate analysis (support vector machine) or univariate analysis (ANOVA)

| Mixed p1000—top 20 | ANOVA—mixed p1000 | Mixed p0.05—top 20 | ANOVA—mixed p0.05 |
|---|---|---|---|
| *LITAF* | *MAL2* | *LITAF* | *TACSTD1* |
| *CLDN8* | *CLDN8* | *SH3YL1* | *SH3YL1* |
| *NDRG1* | *C14orf108* | *NDRG1* | *CA2* |
| *C14orf87* | *C14orf87* | *MYLK* | *C14orf108* |
| *NBEA* | *HLA-G* | *GPR160* | *DKFZP566J2046* |
| *MAL2* | *HLA-C* | *C14orf87* | *C14orf108* |
| *MYLK* | *HLA-B* | *OK/SW-cl.56* | *C14orf114* |
| *APOE* | *HLA-F* | *CA2* | *C14orf87* |
| *APOE* | *embl-id\|M80469* | *MYLK* | *HLA-G* |
| *AKR1C1* | *OK/SW-cl.56* | *APOE* | *HLA-C* |
| *LIMS1* | *BNIP3* | *APOE* | *GDI2* |
| *AKR1C1* | *APOE* | *FER1L3* | *HLA-B* |
| *OK/SW-cl.56* | *CPNE2* | *LIMS1* | *HLA-F* |
| *CPNE2* | *APOE* | *DNASE1L1* | *embl-id\|M80469* |
| *GPR116* | *CUGBP2* | *GPR116* | *OK/SW-cl.56* |
| *BLNK* | *APOE* | *PLEKHA1* | *BNIP3* |
| *embl-id\|M80469* | | *BHLHB2* | *DOCK1* |
| *HSD17B12* | | *CPNE2* | *BNIP3* |
| *CALM1* | | *LOC283177* | *APOE* |
| *HLA-F* | | *STAC2* | *CPNE2* |
| | | | *APOE* |
| | | | *CUGBP2* |
| | | | *APOE* |

25 fresh tumor tissues (10 chromophobe RCC and 15 oncocytoma), including the 18 cases used for microarray.

Conventional semiquantitative RT-PCR was used for initial evaluation and representative results are shown in Fig. 2 (MAL2, AP1M2, and FLJ20171 were tested by quantitative RT-PCR only; see below). These results confirmed the trend of higher expression of *APOE* gene in oncocytoma and all other genes in chromophobe RCC. However, only limited difference was seen between the APOE mRNA levels of oncocytoma and chromophobe RCC, and expression of KRT7, CEL, and TMC4 at levels similar to chromophobe RCC was observed in one or several oncocytomas. In contrast, PRSS8 and PROM2 showed almost no expression in oncocytoma, with strong universal expression in chromophobe RCC. Of interest, normal kidney also showed significant expression of all these genes by RT-PCR (data not shown), indicating that these genes are normally expressed in kidney but were down-regulated in oncocytoma.

Quantitative RT-PCR was then used to evaluate the expression of PRSS8, PROM2, MAL2, AP1M2, and FLJ20171. All five markers showed highly significant difference at their expression levels between chromophobe RCC and oncocytoma ($P < 0.001$), as shown by $\Delta$Ct value distribution in Fig. 3A. Because the ranges of expression were nonoverlapping for all five genes, the diagnosis of all cases could be accurately predicted by the expression of any of the five genes (Fig. 3A).

***Independent validation on formalin-fixed tumor samples.*** RNAs extracted from 15 formalin-fixed, paraffin-embedded tumor tissues (10 chromophobe RCC and 5 oncocytoma) were then analyzed. All cases were different from the ones used in the fresh tissue panel above.

Figure 3B shows the quantitative RT-PCR results. The differential expression of these genes was similarly confirmed as in fresh tissues, with ANOVA *P* values ranging from 0.004 to 0.023. However, the ranges of expression levels were broader than the corresponding values in the fresh sample group. This finding, at least partially due to the higher variation in RNA quality of formalin-fixed tissues, resulted in less effective separation of the two tumor groups. Only *AP1M2* and *FLJ20171* remained capable of separating chromophobe RCC from oncocytomas in this group of 15 cases.

**Table 3.** Differentially expressed genes selected for validation

| Gene name | UniGene | GenBank sequence | Description |
|---|---|---|---|
| *MAL2* | Hs.202083 | NM_052886 | mal, T-cell differentiation protein 2 |
| *AP1M2* | Hs.18894 | NM_005498 | adaptor-related protein complex 1, $\mu$2 subunit |
| *FLJ20171* | Hs.487471 | NM_017697 | RNA binding motif protein 35A (FLJ20171) |
| *PRSS8* | Hs.75799 | NM_002773 | Serine protease 8 (prostasin) |
| *PROM2* | Hs.469313 | NM_144707 | prominin 2 |
| *CLDN8* | Hs.162209 | NM_199328 | claudin 8 |
| *CLDN7* | Hs.513915 | NM_001307 | claudin 7 |
| *CEL* | Hs.533258 | BC042510 | carboxyl ester lipase |
| *KRT7* | Hs.411501 | NM_005556 | keratin 7 |
| *TMC4* | Hs.355126 | NM_144686 | transmembrane channel-like 4 |
| *APOE* | Hs.515465 | NM 000041 | apolipoprotein E |

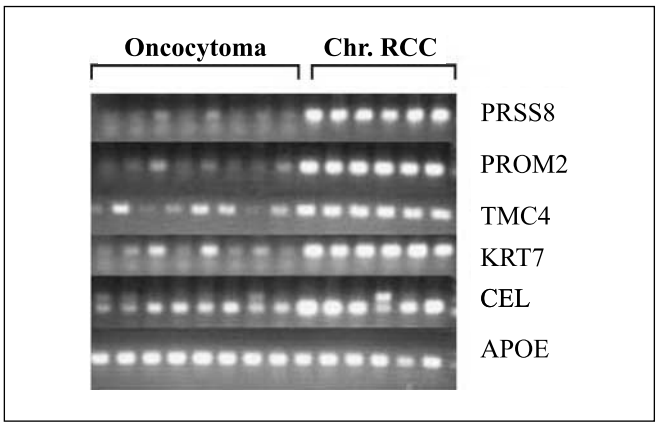


**Fig. 2.** Representative RT-PCR results. PRSS8 and PROM2 showed higher expression in chromophobe RCC, with no or low expression in oncocytoma. In comparison, some oncocytomas showed significant expression of TMC4, KRT7, and CEL. APOE, predicted to have higher expression in oncocytoma, showed only marginal differences between the two groups.

***Immunohistochemical analysis of MAL2 and CLDN8.*** Immunohistochemical analysis was done with antibodies against MAL2 and CLDN8 (Fig. 4). In normal kidney, both antibodies stained distal nephrons, with no or weaker staining in glomeruli and proximal tubules (Fig. 4). By comparing serial sections, these two antibodies appeared to stain the same set of tubules, suggesting coexpression. Both antibodies showed cytoplasmic staining; MAL2 antibody showed a distinctive granular staining pattern with accentuation in the apical cytoplasm, in comparison with the more diffuse staining pattern of CLDN8.

Chromophobe RCC and oncocytoma were then tested. CLDN8 antibody revealed diffuse cytoplasmic staining in both tumor types. In contrast, diffuse MAL2 protein expression was seen in all five chromophobe RCC, but not in oncocytomas (Fig. 4). Of the five oncocytomas, four were negative in most (>99%) cells, with individual positive cells scattered in the tumor. One case, however, showed distinctive clusters of positively stained cells (comprising ~5% of tumor population in total) amidst a negative background, indicating intratumor heterogeneity in expression. Review of the histology showed no distinguishable features in these positive clusters. All MAL2-positive tumors showed cytoplasmic staining pattern. Intriguingly, various distinctive staining patterns were observed in the chromophobe RCC. In three cases, there was accentuation of the cell membranes, with two cases showing a "pericanalicular" or "peritubular" accentuation, as if recapitulating the apical expression pattern of the distal tubule. In another case, diffuse cytoplasmic staining was seen with perinuclear dots noted, suggesting a possible association with organelles, such as Golgi-rough endoplasmic reticulum complex (Fig. 4).

***Functional analysis of differentially expressed genes.*** We asked if the 67 genes from Table 1 could be related and function in one or several common pathways. An EASE analysis (see Materials and Methods) revealed three cellular components as overrepresented in the gene list: tight junction, intracellular junction, and cell junction (*CLDN7*, *CLDN8*, and *TJP3*). The two most significant biological processes were endocytosis and vesicle-mediated transport (*DOCK1*, *AP1M2*, *HOOK2*, and *GDI2*; Table 4). *MAL2*, although not identified in the EASE analysis, was described as an element of basolateral-to-apical transcytosis, and hence is also related to the latter group (26, 29). An Ingenuity pathway analysis confirmed these findings (data not shown). With the exception of *DOCK1* and *GDI2*, which showed higher expression in oncocytomas, all other genes (*CLDN7*, *CLDN8*, *TJP3*, *AP1M2*, and *MAL2*) were preferentially expressed in chromophobe RCC.

## Discussion

Earlier microarray studies on renal tumors found that chromophobe RCC and oncocytoma were highly similar in mRNA expression profile and could not be separated with unsupervised hierarchical clustering (4–8). In our current study, the 18 cases examined clustered into two distinct groups. This difference may be attributed to a few factors. The first is the different gene chips used. It is possible that with >54,000 probe sets and 38,500 genes, the Affymetrix U133 plus 2.0 chip is more powerful in distinguishing these two tumor types than chips used in earlier studies. Another factor was that in all

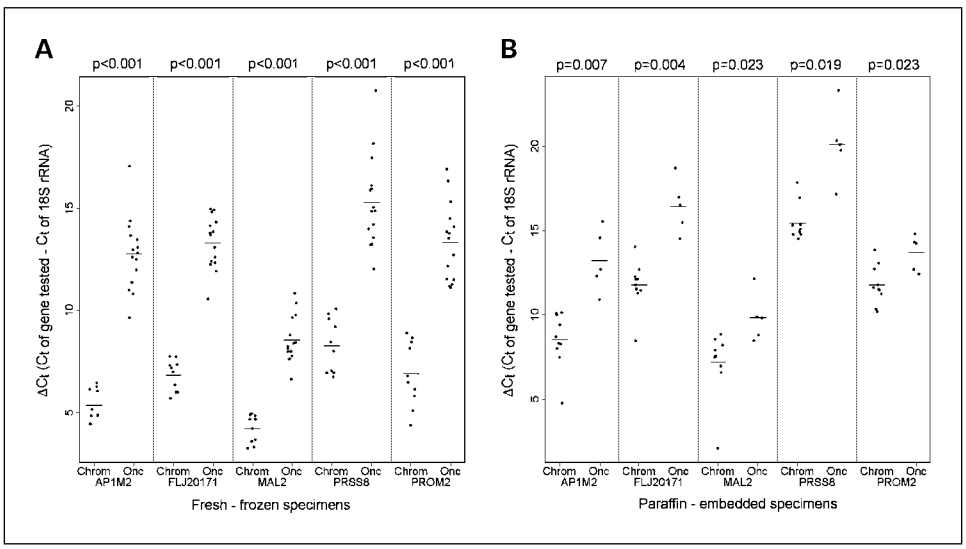


**Fig. 3.** Quantitative RT-PCR analysis of chromophobe RCC (*chrom*) and oncocytoma (*onc*) using RNA extracted from fresh-frozen (*A*) and paraffin-embedded tissues (*B*). Each dot represents a case, with mean values shown as horizontal bars. Lower ΔCt indicates higher mRNA expression, and the calculated *P* values of paired comparisons are shown on top.

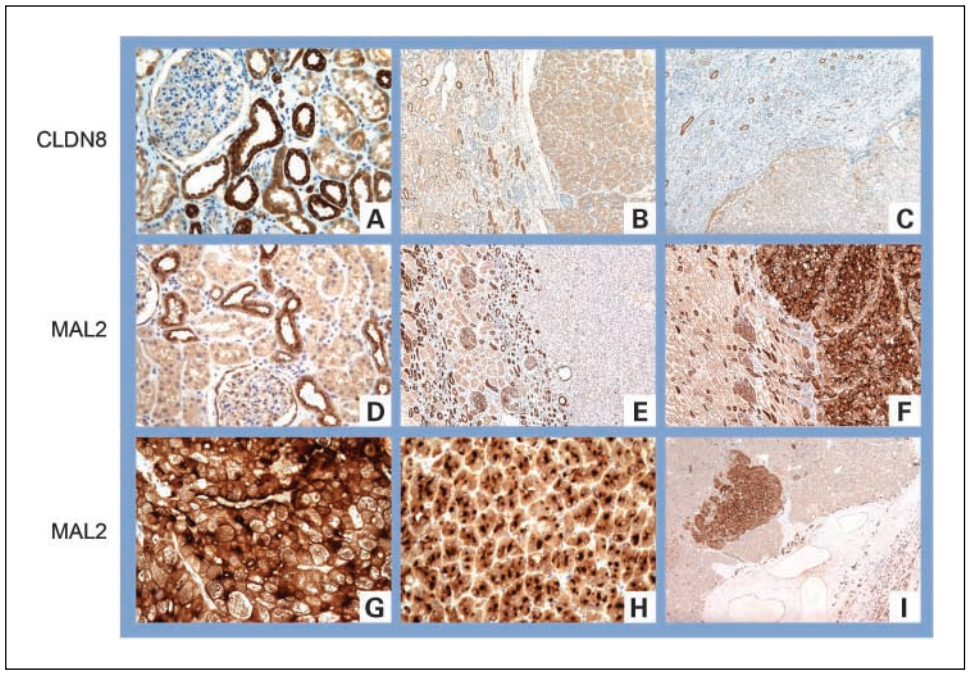


**Fig. 4.** Immunohistochemical analysis of CLDN8 and MAL2. Selective staining of distal nephron in normal kidney was seen (*A* and *D*). Positive CLDN8 staining was also seen in the representative oncocytoma (*B*) and chromophobe RCC (*C*). In contrast, MAL2 stained only chromophobe RCC (*F*) but not most oncocytoma (*E*; see text). MAL2 staining in chromophobe RCC was cytoplasmic, with perimembranous/peritubular accentuation in some cases (*G*) and perinuclear dot pattern in others (*H*). One of five oncocytomas was focally positive for MAL2 (*I*). Magnification, ×400 (*A, D, G*, and *H*); ×200 (*I*); ×100 (other images).

previous studies, only a few cases of chromophobe RCC and oncocytoma were analyzed among a much larger pool of RCC, most cases being of clear cell and papillary types. It is possible that the inclusion of these unrelated tumor subtypes might have obscured the differences between chromophobe RCC and oncocytoma during statistical analysis. The third factor was that only morphologically unequivocal cases of chromophobe RCC and oncocytoma were used for the current study for the purpose of identifying differentially expressed genes, and diagnostically equivocal cases were excluded intentionally. This last factor also means that our finding of distinctive clustering certainly cannot be used as evidence to rule out the possible presence of biologically hybrid tumors, such as those observed in Birt-Hogg-Dube syndrome (30).

Although we successfully separated these two entities by microarray gene profiling, we also found that distinguishing tumor markers could not be easily identified from a gene list (of *P* values and fold changes) alone, and use of multiple statistical variables as well as manual selection was necessary in this selection process. Some of the differentially expressed genes on our gene lists have previously been identified as possible markers, including *CLDN7*, *CLDN8*, *PVALB*, *CK7*, and *MAL2* (16, 29). Our data confirmed that CK7 and PVALB are strongly expressed in chromophobe RCC and low in most oncocytomas. However, we also showed that occasional oncocytomas can have substantial expression of these genes, hence the limitation of CK7 and PVALB as diagnostic markers. This also seemed to be the case for CLDN7 (4) and CDLN8. In contrast, the

**Table 4.** Functional category enrichment analysis (EASE) of the differentially expressed genes

| System | Gene category | EASE score | GenBank accession nos. | Common names |
|---|---|---|---|---|
| GO Biological Process | endocytosis | 0.02606367 | NM_001380; NM_005498; NM_013312 | *DOCK1, AP1M2, HOOK2* |
| GO Biological Process | vesicle-mediated transport | 0.034540003 | D13988; NM_001380; NM_005498; NM_013312 | *GDI2, DOCK1, AP1M2, HOOK2* |
| GO Biological Process | epidermal growth factor receptor signaling pathway | 0.03795122 | NM_017697; NM_024939 | *FLJ20171, FLJ21918* |
| GO Cellular Component | tight junction | 0.005887671 | AL049977; NM_001307; NM_014428 | *CLDN8, CLDN7, TJP3* |
| GO Cellular Component | intercellular junction | 0.027991122 | AL049977; NM_001307; NM_014428 | *CLDN8, CLDN7, TJP3* |
| GO Cellular Component | apicolateral plasma membrane | 0.036015288 | AL049977; NM_001307; NM_014428 | *CLDN8, CLDN7, TJP3* |
| GO Cellular Component | cell junction | 0.037145511 | AL049977; NM_001307; NM_014428 | *CLDN8, CLDN7, TJP3* |
| KEGG pathway | integrin-mediated cell adhesion | 0.160738645 | NM_001380; NM_005186 | *DOCK1, CAPN1* |
| KEGG pathway | cell communication | 0.160738645 | NM_001380; NM_005186 | *DOCK1, CAPN1* |

NOTE: Seven Gene Ontology (GO) categories with *P* values of <0.05 (EASE score) were identified. Two Kyoto Encyclopedia of Genes and Genomes (KEGG) functional pathways were also highlighted.

previously reported stronger expression of MAL2 protein in chromophobe RCC than in oncocytoma (27) was confirmed, paralleling its mRNA differential expression. This suggests that MAL2 antibody is a potentially diagnostic antibody. However, focal MAL2 expression was seen in one of five cases tested, and a larger-scale testing on cases that include eosinophilic chromophobe RCC and hybrid chromophobe RCC/oncocytoma would be necessary to validate the diagnostic usefulness of MAL2.

In addition to the fact that protein expression may not reflect mRNA differential expression, another commonly encountered problem in trying to convert microarray findings to antibody-based immunohistochemical analysis is that genes of interest often encode unknown proteins (e.g., *FLJ20171*) or proteins with no antibodies available for immunohistochemical assays (e.g., PRSS8, PROM2, and AP1M2). Although antibodies may ultimately be available for most biologically interesting molecules, we argue that a potential alternative to immunohistochemistry would be to use the mRNA differential expression, as defined by quantitative RT-PCR, as the assay end point. We have previously shown this approach to be feasible in the differential diagnosis of clear cell RCC, papillary RCC, and chromophobe RCC/oncocytoma (25). We now show that with the five new markers described in this study, most chromophobe RCC and oncocytoma can also be reliably distinguished, even when using formalin-fixed, paraffin-embedded tissue. It is clear that most RCC cases can be classified morphologically, and there might be limited commercial interest in developing an RT-PCR–based assay for RCC. However, similar assays have been developed for diagnostic pathology such as breast cancer prognostication, and we believe that RT-PCR–based assays, in general, could potentially be of value in surgical pathology in the future.

Besides the potential diagnostic usefulness, our findings provide important clues to the biological differences between chromophobe RCC and oncocytoma. Recognizing that the latter is a benign tumor, we anticipated that genes related to malignant transformation or proliferation might be overexpressed in chromophobe RCC. Instead, we found most distinguishing markers with high expression in chromophobe RCC to be genes that encode proteins that represented differentiation markers. By in silico analysis of Expressed Sequence Tags database, many of these genes, including *AP1M2*, *MAL2*, *CLDN7*, and *CLDN8*, had restricted expression in normal tissues, with kidney being one or the main tissue of mRNA expression. Immunohistochemical studies further identified distal nephron cells, the proposed origin of chromophobe RCC and oncocytoma, as the site of expression for CLDN7 (4), MAL2 (27), and CLDN8 (in the present study). Functional analysis revealed that many of these differentially expressed proteins were either tight junction components or proteins involved in apicolateral transcytosis, a finding further supported by the MAL2 subcellular distribution pattern observed by immunohistochemical staining of distal tubules and chromophobe RCC. The lack of expression of these genes in oncocytoma thus implies that the two tumors either are derived from two different cell types or, if from the same cell type, represent two different stages of differentiation, with chromophobe RCC retaining more of the distal nephron markers than oncocytoma.

## Acknowledgments

We thank Lydia Sánchez (Centro Nacional de Investigaciones Oncológicas, Madrid, Spain) for her help on the immunohistochemical analysis of MAL2.

## References


1. Eble JN, Togashi K, Pisani P, et al. Renal cell carcinoma. In: Eble JN, Sauter G, Epstein JI, Sesterhenn IA, editors. WHO classification of tumours: pathology and genetics of tumours of the urinary system and male genital organs. Lyon: IARC Press; 2004. p. 10–32.
2. Amin MB, Tamboli P, Javidan J, et al. Prognostic impact of histologic subtyping of adult renal epithelial neoplasms: an experience of 405 cases. Am J Surg Pathol 2002;26:281–91.
3. Storkel S, Steart PV, Drenckhahn D, et al. The human chromophobe cell renal carcinoma: its probable relation to intercalated cells of the collecting duct. Virchows Arch B Cell Pathol Incl Mol Pathol 1989;56:237–45.
4. Schuetz AN, Yin-Goen Q, Amin MB, et al. Molecular classification of renal tumors by gene expression profiling. J Mol Diagn 2005;7:206–18.
5. Takahashi M, Yang XJ, Sugimura J, et al. Molecular subclassification of kidney tumors and the discovery of new diagnostic markers. Oncogene 2003;22:6810–8.
6. Young AN, Amin MB, Moreno CS, et al. Expression profiling of renal epithelial neoplasms: a method for tumor classification and discovery of diagnostic molecular markers. Am J Pathol 2001;158:1639–51.
7. Higgins JP, Shinghal R, Gill H, et al. Gene expression patterns in renal cell carcinoma assessed by complementary DNA microarray. Am J Pathol 2003;162:925–32.
8. Boer JM, Huber WK, Sultmann H, et al. Identification and classification of differentially expressed genes in renal cell carcinoma by expression profiling on a global human 31,500-element cDNA array. Genome Res 2001;11:1861–70.
9. Speicher MR, Schoell B, du Manoir S, et al. Specific loss of chromosomes 1, 2, 6, 10, 13, 17, and 21 in chromophobe renal cell carcinomas revealed by comparative genomic hybridization. Am J Pathol 1994;145:356–64.
10. Iqbal MA, Akhtar M, Ali MA. Cytogenetic findings in renal cell carcinoma. Hum Pathol 1996;27:949–54.
11. Brunelli M, Eble JN, Zhang S, Martignoni G, Delahunt B, Cheng L. Eosinophilic and classic chromophobe renal cell carcinoma have similar frequent losses of multiple chromosomes from among chromosomes 1, 2, 6, 10, and 17, and this pattern of genetic abnormality is not present in renal oncocytoma. Mod Pathol 2005;18:161–9.
12. Brown JA, Takahashi S, Alcaraz A, et al. Fluorescence *in situ* hybridization analysis of renal oncocytoma reveals frequent loss of chromosomes Y and 1. J Urol 1996;156:31–5.
13. Mazal PR, Exner M, Haitel A, et al. Expression of kidney-specific cadherin distinguishes chromophobe renal cell carcinoma from renal oncocytoma. Hum Pathol 2005;36:22–8.
14. Shen SS, Krishna B, Chirala R, Amato RJ, Truong LD. Kidney-specific cadherin, a specific marker for the distal portion of the nephron and related renal neoplasms. Mod Pathol 2005;18:933–40.
15. Wu SL, Kothari P, Wheeler TM, Reese T, Connelly JH. Cytokeratins 7 and 20 immunoreactivity in chromophobe renal cell carcinomas and renal oncocytomas. Mod Pathol 2002;15:712–7.
16. Abrahams NA, MacLennan GT, Khoury JD, et al. Chromophobe renal cell carcinoma: a comparative study of histological, immunohistochemical and ultrastructural features using high throughput tissue microarray. Histopathology 2004;45:593–602.
17. Young AN, de Oliveira Salles PG, Lim SD, et al. β Defensin-1, parvalbumin, and vimentin: a panel of diagnostic immunohistochemical markers for renal tumors derived from gene expression profiling studies using cDNA microarrays. Am J Surg Pathol 2003;27:199–205.
18. Lin F, Yang W, Betten M, The BT, Yang XJ. The French Kidney Cancer Study Group. Expression of S100 protein in renal cell neoplasms. Hum Pathol 2006;37:462–70.
19. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4:249–64.
20. Benjamini Y, Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc B Met 1995;57:289–300.
21. Cortes C, Vapnik V. Support-Vector Networks. Machine Learning 1995;20:273–97.
22. Pavlidis P, Wapinski I, Noble WS. Support vector machine classification on the web. Bioinformatics 2004;20:586–7.

**23.** Joachims T. Learning to classify text using support vector machines. Dordrecht: Kluwer Academic Publishers; 2001. p. 35–45.
**24.** Mukherjee S, Tamayo P, Roger S, et al. Estimating data set size requirements for classifying DNA microarray data. J Comput Biol 2003;10:119–42.
**25.** Chen YT, Tu JJ, Kao J, Zhou XK, Mazumdar M. Messenger RNA expression ratios among four genes predict subtypes of renal cell carcinoma and distinguish oncocytoma from carcinoma. Clin Cancer Res 2005; 11:6558–66.
**26.** de Marco MC, Martin-Belmonte F, Kremer L, et al. MAL2, a novel raft protein of the MAL family, is an essential component of the machinery for transcytosis in hepatoma HepG2 cells. J Cell Biol 2002;159:37–44.
**27.** Marazuela M, Acevedo A, Garcia-Lopez MA, Adrados M, de Marco MC, Alonso MA. Expression of MAL2, an integral protein component of the machinery for basolateral-to-apical transcytosis, in human epithelia. J Histochem Cytochem 2004;52: 243–52.
**28.** Dennis G, Jr., Sherman BT, Hosack DA, et al. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 2003;4:P3.
**29.** Marazuela M, Martin-Belmonte F, Garcia-Lopez MA, Aranda JF, de Marco MC, Alonso MA. Expression and distribution of MAL2, an essential element of the machinery for basolateral-to-apical transcytosis, in human thyroid epithelial cells. Endocrinology 2004;145:1011–6.
**30.** Pavlovich CP, Walther MM, Eyler RA, et al. Renal tumors in the Birt-Hogg-Dube syndrome. Am J Surg Pathol 2002;26:1542–52.

Clinical Cancer Research

# Gene Expression Profiling Separates Chromophobe Renal Cell Carcinoma from Oncocytoma and Identifies Vesicular Transport and Cell Junction Proteins as Differentially Expressed Genes

Stephen Rohan, Jiangling J. Tu, Jean Kao, et al.

*Clin Cancer Res* 2006;12:6937-6945.