Expression of histone methyltransferases as novel biomarkers for renal cell tumor diagnosis and prognostication

Ana Silvia Pires-Luís1,2, Márcia Vieira-Coimbra1,2,1, Filipa Quintela Vieira1,3, Pedro Costa-Pinheiro1, Paula C Dias2, Luís Antunes4, Francisco Lobo5, Jorge Oliveira5, Céline S Gonçalves5,6,7, Bruno M Costa6,7, Rui Henrique1,2,8,1, and Carmen Jerónimo1,8,1,*

1Cancer Biology and Epigenetics Group – Research Center; Portuguese Oncology Institute – Porto; Porto, Portugal; 2Department of Pathology; Portuguese Oncology Institute – Porto; Porto, Portugal; 3School of Allied Health Sciences (ESTSP); Polytechnic of Porto; Porto, Portugal; 4Department of Epidemiology; Portuguese Oncology Institute – Porto; Porto, Portugal; 5Department of Urology; Portuguese Oncology Institute; Porto; Porto, Portugal; 6Department of Urology; Portuguese Oncology Institute – Porto; Porto, Portugal; 7Life and Health Sciences Research Institute (ICVS); School of Health Sciences; University of Minho; Braga, Portugal; 8ICVS/3B’s – PT Government Associate Laboratory; University of Minho; Braga/Guimarães; Portugal; 9Department of Pathology and Molecular Immunology; Institute of Biomedical Sciences Abel Salazar (ICBAS) – University of Porto; Porto, Portugal;

1Joint first authors; 2Joint senior authors

Keywords: epigenetic biomarkers, histone covalent modifications, histone methyltransferases, NO66, Renal cell tumors, SMYD2, SETD3

Renal cell tumors (RCTs) are the most lethal of the common urological cancers. The widespread use of imaging entailed an increased detection of small renal masses, emphasizing the need for accurate distinction between benign and malignant RCTs, which is critical for adequate therapeutic management. Histone methylation has been implicated in renal tumorigenesis, but its potential clinical value as RCT biomarker remains mostly unexplored. Hence, the main goal of this study was to identify differentially expressed histone methyltransferases (HMTs) and histone demethylases (HDMs) that might prove useful for RCT diagnosis and prognostication, emphasizing the discrimination between oncocyota (a benign tumor) and renal cell carcinoma (RCC), especially the chromophobe subtype (chRCC). We found that the expression levels of 3 genes—SMYD2, SETD3, and NO66—was significantly altered in a set of RCTs, which was further validated in a large independent cohort. Higher expression levels were found in RCTs compared to normal renal tissues (RNTs) and in chRCCs comparatively to oncocytomas. SMYD2 and SETD3 mRNA levels correlated with protein expression assessed by immunohistochemistry. SMYD2 transcript levels discriminated RCTs from RNT, with 82.1% sensitivity and 100% specificity [area under curve (AUC) = 0.959], and distinguished chRCCs from oncocytomas, with 71.0% sensitivity and 73.3% specificity (AUC = 0.784). Low expression levels of SMYD2, SETD3, and NO66 were significantly associated with shorter disease-specific and disease-free survival, especially in patients with non-organ confined tumors. We conclude that expression of selected HMTs and HDMs might constitute novel biomarkers to assist in RCT diagnosis and assessment of tumor aggressiveness.

Introduction

Kidney cancer is the most lethal of the common urological cancers, with 337,860 new cases and 143,369 deaths registered worldwide in 2012.1 In the United States, 63,920 new cases and 13,860 deaths due to kidney cancer were estimated in 2014.2 Renal cell tumors (RCTs), which originate from renal tubule cells, are the most frequent kidney neoplasms, accounting for 85% to 90% of all cases.3 RCTs are a heterogeneous group of neoplasms, comprising several different histological subtypes, each with distinct morphologic, genetic, and clinical features. The four major RCT subtypes include 3 malignant tumors [clear cell renal cell carcinoma (ccRCC, the most common subtype, 75% of all RCT), papillary RCC (pRCC, 10% of all RCT), and chromophobe RCC (cRCC, 5% of all RCT)] and a benign tumor (renal oncocytoma), comprising 3 to 5% of all adult RCT.3 Due to their different clinical aggressiveness, accurate classification is required for appropriate patient management.

The widespread use of imaging techniques has increased the detection of small renal masses, requiring novel tools for accurate diagnosis. Currently, RCT diagnosis relies on histopathological examination of biopsy or surgical specimens, a task that is challenging owing to the overlapping features of some tumor subtypes. One of the most common diagnostic problems is the
differential diagnosis between renal oncocytoma and chRCC, especially its eosinophilic variant.\(^4,5\) Although these tumors share some morphologic and imagiological features, they are biologically distinct, as chRCCs constitute low-grade malignant neoplasms that sometimes behave aggressively and carry a risk of recurrence and metastization, whereas oncocytomas are benign tumors, that might be more conservatively managed.\(^6\) Among RCC, prognostication is also a challenging task. Although pathological stage, nuclear grade, histologic subtype and performance status are commonly used in clinical practice,\(^7\) their ability to accurately predict tumor behavior is limited.\(^8,9\)

Epigenetic alterations, comprising aberrations in DNA methylation patterns, deregulated chromatin machinery, and non-coding RNAs expression, play a critical role in neoplastic transformation, including renal carcinogenesis.\(^10-12\) The modulation of chromatin conformation through covalent posttranslational histone modifications is a fundamental mechanism of gene transcription regulation, which includes methylation, acetylation, phosphorylation, ubiquitylation, and sumoylation of specific residues.\(^13\) Specifically, histone methylation, occurring at the side chains of lysine or arginine residues, is a dynamic process mediated by histone methyltransferases (HMTs) and histone demethylases (HDMs). More than 50 HMTs and HDMs have been identified thus far and, unlike other histone modifications, methylation does not change the charged state of the residues and, therefore, the effect on gene expression is dependent on the residue and its methylation level (mono-, di-, or tri-methylation).\(^14,15\) In RCTs, the deregulation of chromatin machinery has been increasingly acknowledged as an important mechanism of neoplastic transformation (comprehensively reviewed in\(^16\)), highlighting its potential role as diagnostic and prognostic biomarker.

The main goal of this study was to identify HMTs and HDMs that might be used as biomarkers to assist in diagnosis and prognosis of RCT. For that purpose, we screened 87 HMTs and HDMs genes for differential expression between normal renal tissue and RCT, as well as among RCT subtypes, with a particular emphasis on the discrimination of chRCC from oncocytoma. Differentially expressed genes were then validated in a large series of RCTs and renal normal tissues (RNT).

Finally, its clinical usefulness as biomarkers was assessed and compared with standard clinicopathological parameters.

### Results

**Screening of histone methyltransferases and demethylases**

Globally, the analysis of 58 HMTs and 29 HDMs expression levels in 5 chRCCs, 5 oncocytomas and 5 RNTs, disclosed HMTs upregulation in RCTs compared to RNTs. Conversely, HMTs and HDMs were generally downregulated in chRCCs compared to oncocytomas (Fig. S1). However, only SMYD2 (\(P = 0.01\), SETD3 (\(P = 0.005\)), and NO66 (\(P = 0.014\)) showed significantly higher expression levels in RCTs compared to RNTs, and were simultaneously overexpressed in chRCCs in comparison to oncocytomas.

**Validation of selected genes**

Validation of SMYD2, SETD3 and NO66 by RT-qPCR in a series of 160 RCTs and 10 RNTs confirmed that these 3 enzymes were significantly overexpressed in RCTs compared to RNTs (\(P < 0.001\) for SMYD2 and SETD3, \(P = 0.001\) for NO66; Fig. 1A, 1-3). Additionally, expression levels of SETD3 and

---

**Figure 1.** Distribution of expression levels of selected genes. Comparison between RCTs and RNTs for SMYD2 (A1), SETD3 (A2), and NO66 (A3). Benign tumors versus malignant tumors for SMYD2 (B1), SETD3 (B2), and NO66 (B3). Distribution of SMYD2 (C1), SETD3 (C2), and NO66 (C3) expression levels according to renal cell tumor subtype. (**** \(P < 0.0001\); *** \(P < 0.001\); ** \(P < 0.01\)).
NO66 differed significantly between benign and malignant RCTs ($P = 0.003$ and $P = 0.001$, respectively; Fig. 1B, 1-3). Among the 4 RCT subtypes analyzed, chRCC displayed the highest expression levels for the 3 genes, followed by oncocytoma and then by pRCC and ccRCC (Fig. 1C, 1-3). Pairwise comparisons showed statistically significant differences of SMYD2, SETD3, and NO66 expression, in chRCC vs. pRCC ($P < 0.001$, $P < 0.001$, and $P = 0.004$, respectively) and chRCC vs. ccRCC ($P < 0.001$ for all), as well as for SETD3 and NO66 expression in oncocyto vs. pRCC ($P < 0.001$ and $P = 0.004$, respectively) and oncocytoma vs. ccRCC ($P < 0.001$ for both). Additionally, SMYD2 expression levels differed significantly ($P < 0.001$) between chRCC vs. oncocytoma (Fig. 1C, 1-3), in this validation series.

Correlation analysis for SMYD2, SETD3, and NO66 expression in RCC revealed that these 3 genes were significantly co-expressed (SMYD2 – SETD3: $r = 0.759$; SMYD2 – NO66: $r = 0.639$; SETD3 – NO66: $r = 0.741$; $P < 0.001$ for all).

### Association between gene expression and clinicopathological features

Clinical and pathological features of patients included in this study are depicted in Table 1. No significant differences in gender were apparent between RCT patients and controls ($P = 0.524$). In RCTs, NO66 expression levels were significantly higher in females ($P = 0.044$) and SMYD2 expression levels were associated with patient’s age ($P = 0.031$). In malignant tumors, no statistically significant associations were disclosed between SMYD2, SETD3, or NO66 expression levels and pT (Table S1, Fig. S2), nor with pathological stage ($P = 0.692$, $P = 0.724$ and $P = 0.843$, respectively). SMYD2 and SETD3 expression levels were significantly higher in Fuhrman grade 1/2 vs. 3/4 ccRCCs and pRCCs ($P = 0.045$ and $P = 0.021$, respectively; Table S2, Fig. S3). Fuhrman grading was not applied to chRCCs because this grading system does not reflect chRCC clinical aggressiveness and an alternative grading system has been proposed.

### Diagnostic performance of selected genes

The diagnostic performance of SMYD2, SETD3, and NO66 mRNA expression was assessed in 3 different settings, using ROC curve analysis: (i) identification of RCTs vs. renal normal tissue; (ii) discrimination of malignant from benign RCTs; and (iii) distinction of chRCC from oncocytoma (Table 2). Whereas SMYD2 expression levels discriminated RCTs from normal kidney (80.6% sensitivity, 100% specificity, AUC = 0.961; Table 2 and Fig. 2), the performance of all 3 genes in distinction of malignant from benign tumors was modest (highest AUC = 0.671, for NO66) (Table 2). Finally, SMYD2 expression levels could distinguish chRCCs from oncocytomas (AUC = 0.794) with 72.5% sensitivity and 72.5% specificity (Table 2).

### Survival analysis

The median follow-up of RCC patients was 167 months (range: 1-391 months). When considering the 7 years period defined for survival analysis, 10 patients died and 13 developed metastasis. Patients with kidney-confined tumors (Stage I and Stage II) displayed a
significantly higher disease-specific survival (DSS) \((P < 0.001,\) respectively); however, no statistically significant difference was apparent for DFS \((P = 0.055).\) DSS \((P = 0.018),\) but not DFS, was also significantly different among the 3 RCC subtypes. DSS and DFS did not associate with age, gender, or Fuhrman grade for the 3 RCC subtypes in this series. However, considering only ccRCC and pRCC, Fuhrman grade 4 was associated with lower DSS \((P < 0.001)\) and DFS \((P < 0.001).

Focusing on the expression levels of the validated enzymes, survival analysis showed that low \textit{SMYD2}, \textit{SETD3}, and \textit{NO66} expression levels were significantly associated with shorter DSS \((P = 0.012, P = 0.001,\) and \(P = 0.011,\) respectively; Fig. 3A) and DFS \((P < 0.001, P < 0.001\) and \(P = 0.001,\) respectively, Fig. 3B). When evaluating each enzyme’s expression level and pT stage in multivariate analysis (pT and \textit{SMYD2} expression level; pT and \textit{SETD3} expression level; pT and \textit{NO66} expression level), low levels of all 3 genes and pT3 were significantly associated with shorter DSS, and low \textit{SMYD2} expression levels and pT3 were significantly associated with shorter DFS (Table S3). Similar results were depicted for pathological stage (pTNM; data not shown). Conversely, in a model comprising gender, histological subtype, pathological stage, \textit{SMYD2}, \textit{SETD3}, and \textit{NO66} expression levels simultaneously, only patients with tumors not confined to the kidney (Stage III / Stage IV) and that presented metastasis during the follow-up were found to have a significantly increased risk of death due to RCC. Moreover, patients with lower \textit{SMYD2} expression levels and tumors not confined to the kidney (Stage III / Stage IV) presented a significantly increased risk of RCC progression (Table 3).

**Immunohistochemical evaluation of \textit{SMYD2} and \textit{SETD3} expression**

Immunoreactivity for \textit{SMYD2} and \textit{SETD3} was observed in the cytoplasm. In normal renal parenchyma, weak to moderate \textit{SMYD2} and \textit{SETD3} expression was found in tubular epithelial cells. Oncocytomas and chRCC globally displayed higher staining intensity and/or percentage of positive cells, compared to ccRCC and pRCC, following the trend depicted for mRNA expression levels. Indeed, a significant association was documented between mRNA relative expression (RT-qPCR) and protein expression (immunohistochemistry) classified as high and low, both for \textit{SMYD2} \((P = 0.002)\) and \textit{SETD3} \((P = 0.008)\) (Fig. S4). However, \textit{SMYD2} and \textit{SETD3} immunoreactivity was not associated with Fuhrman grade \((P = 0.403 \text{ and } P = 0.110,\) respectively) in ccRCC and pRCC or with pathological stage \((P = 0.636 \text{ and } P = 0.609,\) respectively). DSS \((P = 0.599 \text{ and } P = 0.3\) respectively), and DFS \((P = 0.99 \text{ and } P = 0.192,\) respectively) in RCC.

**RNA expression and survival analysis of renal cell carcinoma patients from TCGA**

Further validation of the results depicted for \textit{SMYD2}, \textit{SETD3} and \textit{NO66} expression was performed in a larger and independent dataset from TCGA, including RNA-seq expression data from 889 RCC patients (533 ccRCC, 290
pRCC, and 66 chRCC) and 129 matched normal samples (n = 72, n = 32, and n = 25, respectively; Table S4). SMYD2 expression was significantly higher ($P < 0.0001$), and SETD3 expression significantly lower ($P < 0.0001$), in RCC than in matched normal samples, whereas no statistically significant difference was disclosed for NO66 expression ($Fig. 4A$). Similarly to the results found in our series, for all 3 genes, statistically significant higher expression levels were found in chRCC compared to pRCC and ccRCC ($Fig. 4B$). Correlation analysis for SMYD2, SETD3, and NO66 expression showed that SMYD2 was significantly co-expressed with SETD3 ($r = 0.24$, $P < 0.0001$) and SETD3 with NO66 ($r = 0.28$, $P < 0.0001$) in RCTs.

SETD3 and NO66 expression levels were significantly lower ($P < 0.0001$) in RCC than in matched normal samples, whereas no statistically significant difference was disclosed for NO66 expression ($Fig. 4A$). Similarly to the results found in our series, for all 3 genes, statistically significant higher expression levels were found in chRCC compared to pRCC and ccRCC ($Fig. 4B$). Correlation analysis for SMYD2, SETD3, and NO66 expression showed that SMYD2 was significantly co-expressed with SETD3 ($r = 0.24$, $P < 0.0001$) and SETD3 with NO66 ($r = 0.28$, $P < 0.0001$) in RCTs.

$$\text{SETD3 and NO66 expression levels were significantly lower (} P < 0.0001\text{}, \text{ in RCC than in matched normal samples, whereas no statistically significant difference was disclosed for NO66 expression (} F\text{ig. 4A}.$$ Similarly to the results found in our series, for all 3 genes, statistically significant higher expression levels were found in chRCC compared to pRCC and ccRCC ($Fig. 4B$). Correlation analysis for SMYD2, SETD3, and NO66 expression showed that SMYD2 was significantly co-expressed with SETD3 ($r = 0.24$, $P < 0.0001$) and SETD3 with NO66 ($r = 0.28$, $P < 0.0001$) in RCTs.

$$\text{SETD3 and NO66 expression levels were significantly lower (} P < 0.0001\text{, in RCC than in matched normal samples, whereas no statistically significant difference was disclosed for NO66 expression (} F\text{ig. 4A}.$$ Similarly to the results found in our series, for all 3 genes, statistically significant higher expression levels were found in chRCC compared to pRCC and ccRCC ($Fig. 4B$). Correlation analysis for SMYD2, SETD3, and NO66 expression showed that SMYD2 was significantly co-expressed with SETD3 ($r = 0.24$, $P < 0.0001$) and SETD3 with NO66 ($r = 0.28$, $P < 0.0001$) in RCTs.

| Table 3. Prognostic factors for Renal Cell Carcinoma obtained by Cox regression analysis |
|---------------------------------------------|----------------|------------------|
| Prognostic Factor                          | Univariate Analysis          | Multivariate Analysis          |
|                                            | HR 95% CI for HR P value* | HR 95% CI for HR P value* |
| Disease Specific Survival                  |                              |                              |
| Pathological Stage                         | 7.905 2.043-30.585 0.003    | 4.937 1.255-19.428 0.022     |
| Metastasis (during follow-up)              | 35.88 7.599-169.36 <0.001   | 27.378 5.702-131.459 <0.001  |
| SMYD2 expression level                     | 4.43 1.25-15.73 0.021       | 0.614                         |
| SETD3 expression level                     | 7.02 1.81-27.21 0.005       | 0.819                         |
| NO66 expression level                      | 4.50 1.27-15.99 0.02        | 0.705                         |
| Histological Subtype                       |                              |                              |
| chRCC                                      |                              |                              |
| pRCC                                       | 2.81 0.94-8.38 0.064        | 3.31 1.105-9.887 0.032       |
| ccRCC                                      | 5.30 1.73-16.23 0.003       | 0.233                         |
| Disease Free Survival                      |                              |                              |
| Pathological Stage                         | 2.81 0.94-8.38 0.064        | 3.31 1.105-9.887 0.032       |
| Metastasis (during follow-up)              | 4.43 1.25-15.73 0.021       | 0.614                         |
| SMYD2 expression level                     | 7.02 1.81-27.21 0.005       | 0.819                         |
| SETD3 expression level                     | 4.50 1.27-15.99 0.02        | 0.705                         |
| Histological Subtype                       |                              |                              |
| chRCC                                      |                              |                              |
| pRCC                                       | 2.81 0.94-8.38 0.064        | 3.31 1.105-9.887 0.032       |
| ccRCC                                      | 5.30 1.73-16.23 0.003       | 0.233                         |
| Gender                                     |                              |                              |
| c                                           | 2.81 0.94-8.38 0.064        | 3.31 1.105-9.887 0.032       |
| HR: Hazard Ratio; CI: Confidence Interval, n.s.: not significant. *Cox regression p value; significant when $P < 0.05$.  

Reference group: high expression level.

Reference group: Absence of metastasis.

Reference group: Male.

www.tandfonline.com 1037Epigenetics

$F\text{igure 4. Distribution of expression levels of selected genes from the TCGA data set (http://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm. Accessed 2015). Comparison between RCCs and RNTs for SMYD2 (A1), SETD3 (A2), and NO66 (A3); Distribution of SMYD2 (B1), SETD3 (B2), and NO66 (B3) expression levels according to renal cell carcinoma subtype. (****} P < 0.0001; *** P < 0.001; ** P < 0.01).$
The median follow-up of RCC patients from TCGA was 29.4 months (range: 0-195 months), in which 212 patients died and 92 developed metastasis. A significantly higher risk of death due to RCC was detected in ccRCC patients with non-organ confined tumors (Stage III and Stage IV), lower SETD3 expression, and pRCC (compared to chRCC), in multivariate analysis ($P < 0.0001$; Table 4). Interestingly, Stage III / Stage IV tumors and lower NO66 expression levels associated with shorter time to disease progression in multivariate analysis ($P < 0.001$ and $P = 0.018$, respectively) (Table 4).

When analyzing each subtype separately, multivariate analysis revealed that Stage III / Stage IV was associated with shorter DSS in all subtypes and shorter DFS in ccRCC and pRCC; occurrence of metastasis during follow-up with shorter DSS in ccRCC and pRCC; low SETD3 expression levels with shorter DSS in ccRCC and pRCC and with shorter DFS in pRCC; and low NO66 expression levels with shorter DSS in pRCC (Table S5). DFS analysis was not performed to chRCC due to insufficient available data in the TGCA database.

**Discussion**

The incidental diagnosis of small renal masses has increased in recent years, and current imaging techniques, even when complemented with biopsy, are limited in discriminating benign from malignant entities, leading to surgical intervention in cases in which it might be spared and the patient be managed conservatively. Epigenetic-based biomarkers, including histone posttranslational modifications and chromatin modulators, hold the promise to assist in diagnosis through the discrimination between normal and neoplastic tissue or between benign and malignant tumors, helping in prognostication and stratifying patients according to risk of disease progression. Evidence of widespread deregulation of chromatin status in RCTs has been accumulating, and several defects in epigenetic enzymes, including those responsible for chromatin packaging, histone modifications, and chromatin remodeling, have been reported.$^{18,19}$

In this study, we focused on altered histone methylation patterns and screened the expression of HMTs and HDMs genes in RNT, oncocytomas and chRCCs to determine whether their expression levels might be used as diagnostic biomarkers to discriminate normal from neoplastic renal tissue, as well as benign (oncocytoma) from malignant RCTs (especially chRCC). For most of the genes assessed, differences in expression levels between normal and neoplastic samples were in line with previous reports on their role in cancer. Thus, $SETD2$, a putative tumor suppressor in ccRCC,$^{20}$ was mostly downregulated in RCT compared to RNTs, and a similar trend was apparent for $KDM6A$, previously reported to be overexpressed in RCTs,$^{21}$ was found to be the second most upregulated...
gene in RCTs in our analysis. A divergent behavior was, however, observed for EZH2, previously shown to be upregulated in ccRCCs compared to adjacent normal tissues, whereas we found EZH2 downregulation in RCTs. Notwithstanding this latter result, which might be due to the small number of cases or the main histological subtype (chRCC) used for array analysis in our study, and/or the use of adjacent morphologically normal renal tissue as control (which we have found to harbor epigenetic alterations) in the aforementioned studies, the overall results of the array and the agreement found for the 3 genes validated in a series of 160 primary tumors, as well as in GEO database (http://www.ncbi.nlm.nih.gov/gds) published array data, argue in favor of the validity of our strategy for discovery of HDMs and HMTs as RCT biomarkers.

We found that 2 histone methyltransferases, SMYD2 and SETD3, and a histone demethylase, NO66, were significantly upregulated in RCTs compared to RNTs, and in chRCC compared to oncocytoma, in the screening array. SMYD2 targets lysines 4 and 36 of histone H3, as well as non-histone proteins including p53, RB1, HSP90, and PARP-1. It plays an important role in muscle function, mediating the methylation of Hsp90, which stabilizes the sarcomeric region, and in the early stages of embryonic differentiation. In cancer, SMYD2 is overexpressed in bladder cancer and leukemia, as well as in esophageal squamous cell carcinoma and gastric carcinoma. SETD3, which also methylates lysines 4 and 36 of histone H3, promotes differentiation of muscle cells and is overexpressed in lymphomas, displaying oncogenic potential. NO66, which specifically targets lysines 4 and 36 of histone H3, is involved in bone differentiation (osteoblast differentiation and bone formation), inducing chromatin repression through histone demethylation during osteoblast differentiation.

Our screening data are in line with “GEO Data sets” publicly available expression array data (http://www.ncbi.nlm.nih.gov/gds) of 18 cases, comprising 9 chRCC and 9 oncocytomas, which also revealed a higher expression of SMYD2, SETD3, and NO66 in chRCC compared to oncocytoma (Table S6). Importantly, this pattern of expression was also retained in our validation series: chRCC and oncocytoma displayed higher expression levels than ccRCC and pRCC. This finding might denote the common cellular origin of ccRCC and pRCC, on the one hand, and of chRCC and oncocytoma, on the other, as previously suggested. Indeed, expression levels of these 3 enzymes were significantly different between chRCC and pRCC, as well as between chRCC and ccRCC, both in our validation series and in the TCGA database. Additionally, SMYD2 expression levels differed significantly between chRCCs and oncocytomas, indicating a potential for discriminating among these 2 RCT subtypes, which frequently display overlapping and confounding morphological features that might impair differential diagnosis, especially in small biopsies. Importantly, mutational analysis data accessible in the COSMIC database (http://cancer.sanger.ac.uk/cosmic) revealed that, in RCC, SMYD2, and SETD3 mutations are exceedingly rare [mutation rate of 0.22% (3/1345) and 0.32% (4/1253), respectively] and were not reported for NO66. Thus, it seems unlikely that the differences in gene expression we observed might be due to genetic alterations.

We also found significantly higher expression levels of SMYD2, SETD3, and NO66 in RCTs than in RNTs in the validation series. These results are similar to those observed in the TCGA dataset for SMYD2 expression, keeping in mind that in our validation series the normal tissue was collected from non-RCT patients, whereas in TCGA database morphologically matched normal renal tissue from RCC patients was analyzed. This difference in normal samples may account for the higher SETD3 expression in normal tissue than in RCC noted in the TCGA data set. Moreover, among the 3 validated genes, SMYD2 expression levels also displayed the best diagnostic performance for distinction between RCTs and normal renal tissue, thus increasing the spectrum of scenarios in which it may assist in diagnosis and classification of suspicious renal lesions. When compared to other molecular techniques, FISH was reported to allow the identification of RCT subtypes, specially to distinguish chRCC from oncocytoma, but the overlapping genetic alterations mainly between chRCC and oncocytoma might hamper differential diagnosis. A molecular algorithm based on qPCR gene expression correctly identified the RCC subtype in 83.3% of cases, and CGH allowed for the correct diagnosis in 93.5% of ccRCC, 100% of pRCC, 61.5% of chRCC, and 14.3% of oncocytomas, although in both studies less than 80% cases were analyzed.

Interestingly, SMYD2 and SETD3 differential expression at transcript level was also apparent at protein level, as assessed by immunohistochemistry, and a statistically significant correlation between the 2 parameters was found. However, the wide variation in immunoexpression within RCT subtype and across subtypes precludes its use as an ancillary tool for histopathological evaluation. Notwithstanding, owing to the increasing availability and use of molecular techniques in diagnostic pathology, the assessment of mRNA expression in tissue samples or in fine-needle aspirates of suspicious lesions is within the reach of many molecular pathology laboratories.

Besides its potential role as diagnostic biomarkers, we aimed also at characterizing the prognostic value of HMTs and HDMs expression in RCCs. Interestingly, low SMYD2, SETD3, and NO66 expression levels associated with worse disease-specific survival and disease-free survival, in univariate analysis. The prognostic value of SMYD2 overexpression has been reported in leukemia, esophageal squamous cell carcinoma, and gastric carcinoma, contrary to our findings in RCC. Besides differences in tumor model, it should be emphasized that survival analyses in our series were mostly influenced by pRCC and ccRCC, which displayed the lowest expression levels for the 3 validated genes, and that are acknowledged as the most aggressive RCC subtypes. Furthermore, in multivariate analysis assessing the expression level of each gene, pT, histological subtype, and gender, statistical significance was retained for pT and expression level for the 3 enzymes, both for DSS and DFS, with low expression level and pT3 associating with worse prognosis, thus indicating that each gene independently adds prognostic information to pT stage. When combining the expression levels of the 3 genes with pathological stage, occurrence of metastasis during follow-up, histological subtype, and gender, for our series and for the
TCGA dataset separately, Stage III / Stage IV and the development of metastasis associated with shorter DSS in both cohorts. Furthermore, low SETD3 expression and ccRCC and pRCC (compared to chRCC) subtypes also associated with decreased DSS, but only in the TCGA data set. A shorter DFS was associated with low SMYD2 expression levels and Stage III / Stage IV in our series, and low NO66 and Stage III / Stage IV in the TCGA dataset. These differences might be due to (i) cohort size, (ii) dissimilar proportions of the 3 histological subtypes in the 2 cohorts, or (iii) differences in period of follow-up. Since the number of events, both cancer-specific deaths and death of metastasis, were scarce in our series, mainly in the ccRCC (subtypes 2 and 4, respectively) and chRCC (subtypes 1 and 2, respectively), survival analysis was not performed separately for each subtype. To overcome this limitation, TCGA data set was further explored. Interestingly, when analyzing each subtype separately in multivariate analysis, SETD3 was associated with shorter DSS in ccRCC and shorter DSS and DFS in pRCC, underlining the potential clinical value of SETD3 as a prognostic biomarker.

To the best of our knowledge, only a few studies have addressed the role of SMYD2, SETD3, and NO66 in tumorigenesis, as previously described, although none has included RCTs. The knowledge of the specific role of SMYD2, SETD3, and NO66 in renal carcinogenesis, uncovering the participation of these enzymes in the neoplastic transformation of renal epithelium and in RCC progression, might additionally provide a stronger biological rational for its use as clinically useful diagnostic and prognostic biomarkers.

Patients and Methods

Patients and sample collection
A total of 160 RCTs, comprising ccRCCs, pRCCs, chRCCs, and oncocytomas (40 cases of each type), were prospectively collected from patients consecutively diagnosed and submitted to nephrectomy at the Portuguese Oncology Institute – Porto, between 2001 and 2014. As controls, 10 renal normal tissue (RNT) samples were collected from morphologic normal kidneys of patients subjected to nephrectomy due to upper urinary tract urothelial carcinoma. All specimens were immediately frozen after surgery and stored at −80°C. Sampling of more than 70% of malignant cell was confirmed by 2 slides stained with hematoxylin and eosin (H&E) taken before and after frozen section collection for RNA extraction. Routine histological slides from formalin-fixed paraffin-embedded (FFPE) tissue of the same surgical specimens were assessed for diagnosis, TNM staging, and Fuhrman grading. Relevant clinical data was also collected from clinical charts. This study was approved by the Institutional Review Board of Portuguese Oncology Institute – Porto [Comissão de Ética para a Saúde-(CES-IPOFG-EPE 518/10)].

RNA extraction
Samples were suspended in TRIzol® reagent (Invitrogen™, Cat.#15596018) and chloroform (Merk Millipore, Cat.#MCX10601) was added to the lysed cells. Total RNA was purified using the Ambion® PureLink RNA Mini Kit (Invitrogen™, Cat.#12183025), according to manufacturer recommendations. RNA concentrations and purity ratios were determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies) and RNA quality was verified by electrophoresis.

Screening of histone methyltransferases and demethylases
A total of 10 RCT (5 oncocytomas and 5 chRCCs) and 5 RNT samples were treated with Ambion® TURBO DNA-free™ kit (Invitrogen, Cat.#1907) to remove any DNA contamination, and then 1 μg of total RNA was reversely transcribed using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems®, Cat.#4368814) according to manufacturer instructions. Expression of histone methylation enzymes (58 HMTs and 29 HDMs), was assessed by RT-qPCR using custom made TaqMan® Array 96-Well expression Plates (Applied Biosystems®, Cat.#4391528) in an ABI-7500 Real-Time PCR system (Applied Biosystems®, Cat.#4351105).

Each gene was run in triplicate and the amount of mRNA was normalized to Glucuronidase β (GUSB) and Human 18S rRNA (18S) reference genes. Comparative CT method was used to determine the fold-difference in gene expression between RNT and RCT, as well as between chRCC and oncocytoma. Genes that reached statistically significant differences in expression levels between these groups, and displayed higher or lower expression levels in RCT vs. RNT or in chRCC vs. oncocytoma, were selected for further analysis.

Validation of selected enzymes
Candidate genes’ mRNA levels were evaluated in a larger series of 160 RCTs, including 40 ccRCCs, 40 pRCCs, 40 chRCCs, 40 oncocytomas, and 10 RNTs. A total of 300ng was reversely transcribed and amplified using TransPlex®Whole Transciptome Amplification Kit (Sigma-Aldrich®, Cat.#WT1) purified with QIAquick PCR Purification Kit (QIA-GEN, Cat.#28106), and mRNA levels were evaluated using TaqMan® Gene Expression Assays [Applied Biosystems®, Cat.# Hs00220210 m1 (SMYD2), Hs00260120 m1 (SETD3), Hs02743012 s1 (NO66), Hs99999908 m1 (GUSB), Hs99999901 s1 (18S)], according to manufacturer’s instructions. For each sample, expression levels were normalized using 2 internal reference genes, GUSB and 18S, according to the formula: target gene relative expression = [GUSB expression level + 18S expression level] / 2. Each plate included multiple non-template controls and serial dilutions of a cDNA Human Reference Total RNA (Agilent Technologies, Cat.#750500) to construct a standard curve.

Immunohistochemistry
A representative slide from 120 RCT cases of the validation cohort (30 cases available from each subtype) was selected and 4 μm sections from FFPE tissue were obtained. Briefly, after deparaffinization antigen retrieval was performed by heating (20 min) in an antigen unmasking solution (Vector Laboratories, Cat.#H3300), endogenous peroxidase activity was neutralized with
0.6% hydrogen peroxide (Merk Millipore, Cat.#107298) for 20 minutes. Then, protein detection was performed using the Novolink™ Max Polymer Detection System (Leica Biosystems, Cat. #RE7260-K), according to manufacturer instructions. Slides were incubated in a humid chamber with rabbit polyclonal antibodies, specific for SET and MYND domain containing 2 (SMYD2) (Sigma Aldrich®, Cat.#HPA029023) in a 1:250 dilution (4°C, overnight) and specific for SET domain containing 3 (SETD3) (Novus Biologicals®, Cat.#NBP-88416) in a 1:200 dilution (room temperature, 1 hr). For NO66 several primary antibodies were tested but none was found to provide reliable results.

All washings were performed with Tris buffered saline with Tween® 20 (TBS-T) (Sigma-Aldrich®, Cat.#T9039). To unveil antigen-antibody reaction binding result, slides were incubated for 7 minutes, in the dark, in a 0.05% (m/v) 3, 3'-diaminobenzidine (DAB) solution (Sigma-Aldrich®, Cat.#D7304) in phosphate-buffered saline (PBS) (Biochrom Ltd., Cat.#L1835). Then, slides were counterstained with hematoxylin (Merk Millipore, Cat. #105174), dehydrated and diaphanized.

Slides were evaluated by 2 pathologists for SMYD2 and SETD3 immunoexpression and classified using a semi-quantitative scale for both staining intensity (0 – no staining; 1 – intensity lower than normal kidney; 2 – intensity equal to normal kidney; 3 – intensity higher than normal kidney) and percentage of positive cells (0 (0 – < 10%; 1 – 10-33%; 2 – 33-67%; 3 – >67%), in each tumor. Staining intensity and percentage of positive cell scores were combined (Score S = staining intensity x percentage of positive cells) to assign a composite score in each tumor, which was then stratified into low expression (S < 4) and high expression (S ≥ 4) groups, which basically correspond to RCTs with less than 33% stained cells or staining intensity lower than normal kidney, and RCTs with at least 33% stained cells with an intensity equal or higher than normal kidney.

TCGA dataset analysis in renal cell carcinoma patients

The Cancer Genome Atlas (TCGA) was used to obtain data on SMYD2, SETD3 and NO66 expression and clinical information, when available, from renal cell carcinoma patients and matched normal tissue samples.52 All expression data from samples hybridized by the University of North Carolina, Lineberger Comprehensive Cancer Center, using Illumina HiSeq 2000 RNA Sequencing version 2 analysis, were downloaded from TCGA data matrix (http://tcga-data.nci.nih.gov/tcga/tcgaDownload.jsp).52 This data set included 533 ccRCC and 72 matched normal patient samples, 290 pRCC and 32 matched normal patient samples, and 66 chRCC and 25 matched normal patient samples. To prevent duplicates, when there was more than one portion per patient, median values were used. The provided value was pre-processed and normalized according to “level 3” specifications of TCGA (see http://cancergenome.nih.gov/dataportal/ for details). Clinical data of each patient was provided by the Biospecimen Core Resources (BCRs). This data is available for download through TCGA data matrix (http://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm).52

Statistical analysis

Chi-square, Fisher’s exact test and, non-parametric tests were used to ascertain the statistical significance of differences among groups of samples, namely Kruskal-Wallis test (KW) for multiple comparisons and Mann-Whitney U test (MW) for pair-wise comparisons, as appropriate. Spearman’s test was carried out to ascertain correlations between age and HMTs or HDMs expression levels, as well as between HMTs and HDMs mRNA expression levels.

Receiver operator characteristics (ROC) curves were constructed to assess the diagnostic performance of biomarkers, by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity), and the area under the curve (AUC) was calculated. Sensitivity, specificity, positive and negative predictive values, and accuracy were calculated based on cutoff values based on ROC curve analysis, prioritizing specificity and then sensitivity.

For survival analysis, a 7-year follow-up time was considered, both for disease-specific survival (DSS) and disease free survival (DFS), as patient recruitment occurred during 13 years. Prognostic significance of standard clinicopathological variables (histological subtype, pathological stage, Fuhrman grade, age, gender) and as well as of HMTs and HDMs expression levels, was assessed by constructing disease-specific and disease-free survival curves using the Kaplan-Meier method, with log-rank test and Cox-regression analysis for each variable (univariate test). For this purpose, expression levels of SMYD2, SETD3, and NO66 were classified as low or high based on the 25th percentile expression value of each gene. Similarly, SMYD2 and SETD3 immunoreactivity was classified as low or high according to Score S. A Cox-regression model using Forward Stepwise (conditional) test comprising the different variables (multivariate test) was also performed, including the 160 RCC patients, both for disease-specific (DSS) and disease-free (DFS) survival. A similar Cox-regression analysis (univariate and multivariate) was performed for patients from the TCGA dataset, and expression levels were classified as low or high based on the 25th percentile expression value of each gene, too.

Statistical significance was set at P < 0.05. Bonferroni’s correction was applied for pairwise comparisons following multiple groups’ analyses. Statistical analysis was performed using SPSS software for Windows, version 22.0 (IBM-SPSS Inc.), and graphs were built using GraphPad Prism 6.0 software for Windows (GraphPad Software Inc.).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Funding

This study was funded by research grants from Research Center of Portuguese Oncology Institute – Porto (CI-IPOP 4-2012) and European Community’s Seventh Framework Program – Grant number FP7-HEALTH-F5-2009-241783. ASP-L and
Supplemental Material

Supplemental data for this article can be accessed on the publisher’s website.


43. Sinha KM, Yasuda H, Coombrum MM, Dent SY, de Crombrugghe B. Regulation of the osteoblast-specific transcription factor Osterix by NO66, a Junmoji family histone demethylase. EMBO J 2010; 29:68-79; PMID:19927124; http://dx.doi.org/10.1038/emboj.2009.332


