Multilevel Whole-Genome Analysis Reveals Candidate Biomarkers in Clear Cell Renal Cell Carcinoma

Andrew H. Girgis, Vladimir V. Iakovlev, Ben Beheshti, Jane Bayani, Jeremy A. Squire, Anna Bui, Marina Mankaruos, Youssef Yousef, Bishoy Khalil, Heba Khella, Maria Pasic, and George M. Yousef

Abstract

Renal cell carcinoma (RCC) is the most common neoplasm of the kidney. We conducted an integrated analysis of copy number, gene expression (mRNA and miRNA), protein expression, and methylation changes in clear cell renal cell carcinoma (ccRCC). We used a stepwise approach to identify the most significant copy number aberrations (CNA) and identified regions of peak and broad copy number gain and loss, including peak gains (3q21, 5q32, 5q34-q35, 7p11, 7q21, 8q24, 11q13, and 12q14) and deletions (1p36, 2q34-q37, 3p25, 4q33-q35, 6q23-q27, and 9p21). These regions harbor novel tumor-related genes and miRNAs not previously reported in renal cell carcinoma. Integration of genome-wide expression data and gene set enrichment analysis revealed 75 gene sets significantly altered in tumors with CNAs compared with tumors without aberration. We also identified genes located in peak CNAs with concordant methylation changes (hypomethylated in copy number gains such as STC2 and CCND1 and hypermethylated in deletions such as CLCNKB, VHL, and CDKN2A/2B). For other genes, such as CA9, expression represents the net outcome of opposing forces (deletion and hypomethylation) that also significantly influence patient survival. We also validated the prognostic value of miRNA let-7i in RCCs. miR-138, located in chromosome 3p deletion, was also found to have suppressive effects on tumor proliferation and migration abilities. Our findings provide a significant advance in the delineation of the ccRCC genome by better defining the impact of CNAs in conjunction with methylation changes on the expression of cancer-related genes, miRNAs, and proteins and their influence on patient survival.

Introduction

Renal cell carcinoma (RCC) is the most common neoplasm of the adult kidney, and the most lethal genitourinary cancer with more than 40% mortality (1). The incidence of RCC has been increasing, and despite advances in early detection and treatment, the rate of mortality has not changed significantly over the last decades (2). The disease is histopathologically heterogeneous, comprising several subtypes of which the most common (~75%) is the clear cell subtype (ccRCC). The molecular heterogeneity of ccRCC makes gauging clinical outcome and treatment response challenging. Delineating the pathogenesis of ccRCCs by investigating the genetic and epigenetic changes and their effects on key molecules and their respective biologic pathways is of crucial importance for the improvement of current diagnostics, prognostics, and drug development (3).

Drugs that target downstream genes of the pVHL/HIF pathway, including tyrosine kinase and mTOR inhibitors, are used to treat metastatic ccRCCs, albeit with modest improvements in survival (4). The development of new therapies targeting the molecular pathways involved in ccRCCs holds the promise for individualized and highly responsive therapy options, thereby marking the era of personalized medicine (5).

Previous insights into the complexity of the ccRCC genome have revealed frequent deletion of the 3p arm harboring the VHL tumor suppressor, often associated with gains of 5q harboring a number of proposed oncogenes (6). Recent studies using high-resolution microarrays have identified additional copy number aberrations (CNA) at lesser frequency (7–12). Previous reports suggested CNAs in ccRCCs to be dynamically related to clinical parameters, such as associations of 4p, 9p, and 14q deletion and 7q, 8q, 20q gains with higher stage, grade, and/or worse prognosis (8, 10, 13–16). In addition, 1q, 12q, and 20q gains and deletions of 9p have been associated with metastatic risk (16). Moreover, ccRCC molecular subtypes have been recently proposed on the basis of gene expression profiles (6, 17, 18).

However, the influence of CNAs on the expression of cancer-related genes and their impact on biologic pathways is largely unknown in ccRCCs. Also, the coordinated interplay of CNAs...
with alternative regulatory mechanisms (such as methylation and mutation) is still poorly understood. The main exception is VHL biallelic inactivation in the majority of sporadic ccRCC cases. Likewise, the integration of CNAs with noncoding miRNAs has been largely overlooked, despite the dynamic impact of miRNAs in ccRCCs (19).

Currently, molecular profiling approaches have allowed for the global analysis of diverse classes of molecules. They have led not only to faster biomarker discovery but also allowed for better understanding of the cross-talk or biologic interactions that contribute to the development of malignancy (20). By integrating genome-wide CNAs, gene, miRNA and protein expression changes, and CpG island methylation, pertinent insight into the interplay of molecular regulatory mechanisms can be observed, thereby connecting pieces of the puzzle to provide a more comprehensive picture of the ccRCC genome.

Materials and Methods

Sample acquisition

Frozen tumor specimens for primary analysis were obtained through the Ontario Tumour Bank and from St. Michael’s Hospital, Toronto, ON, Canada. Archival formalin-fixed, paraffin-embedded (FFPE) tumor and matched normal specimens for validation analyses were obtained from St. Michael’s Hospital. Histopathologic diagnoses were confirmed by 2 independent pathologists. The study was approved by the Research Ethics Board of St. Michael’s Hospital.

DNA extraction and array comparative genomic hybridization

DNA was isolated from frozen tumor specimens from 10 cases of ccRCCs using the Qiagen DNeasy Blood & Tissue Kit and labeled using the Agilent Genomic DNA Enzymatic Labeling Kit following manufacturers’ protocols (version 5.0). The samples were hybridized to the Agilent Human Genome 244K 60-mer oligonucleotide CGH-arrays. Details of the hybridization process can be found in the Supplementary Methods.

Discovery set

Using our array comparative genomic hybridization (aCGH) data, we used a multistep filtration approach to identify significant CNAs in ccRCCs as described below. We validated this approach using publically available datasets mentioned below. To enhance the strength of the analysis, our aCGH data were combined with the public data to form the discovery set. The discovery set included 154 ccRCC tumors with high-resolution aCGH data [oligonucleotide and single-nucleotide polymorphism (SNP)-oligonucleotide arrays] from our experimental and public repository sources as described in the works of Beroukhim and colleagues, Dondeti and colleagues, and Gordan and colleagues (refs. 6, 9, 18; Gene Expression Omnibus accession nos. GSE14994, GSE27852, GSE13282). Matched gene expression data for 59 ccRCCs cases as described in the work of Beroukhim and colleagues (ref. 9; GSE14994), as well as miRNA and protein expression data from studies previously published by our group were integrated with the copy number analysis (ref. 19; GSE23085) and (21).

Genome-wide copy number analysis

The flowchart of our analysis is shown in Fig. 1. We conducted a multistep analysis to detect significant copy number changes: (i) Microarray data were PLIER-normalized against the Phase II 270 HapMap samples of The International HapMap Consortium (22). (ii) The Circular Binary Segmentation algorithm (23) was applied to generate copy number segments. (iii) Frequent copy number variations in the general population as per the Database of Genomic Variants (The Centre for Applied Genomics, Toronto, ON, Canada) and array-specific artifacts shared
between tumor and normal tissue were removed. We used the term “copy number aberrations” to indicate gains or losses of specific chromosomal segments in cancer, whereas “copy number variations” was used to indicate the copy number differences in normal human DNA. (iv) Significant peak regions of gain and loss were identified using the GISTIC method (24) using the Java-improved implementation of the algorithm, JISTIC (25). Thresholds of CN > 2.3 for gains, CN < 1.7 for deletions, and “2.3 < CN > 1.7” for diploid were used as previously described (6) and a q < 0.25 was selected to define significant CNAs. Broad regions of gains and losses were defined as segments that meet the thresholds and span whole chromosomes or more than half of a chromosomal arm. Normalization and segmentation were conducted using the Agilent GeneSpring GX 11 software with default settings.

**Genome-wide mRNA expression analysis**

Gene expression data of ccRCCs and normal kidney as described in the work of Beroukhim and colleagues (9) were normalized using the Robust Multi-array Average method (26). Gene set enrichment analysis (GSEA) was conducted to assess the distribution of underexpressed and overexpressed sets of genes in relation to their genomic location and identify effects of copy number losses or gains on mRNA expression. Enrichment of gene sets was based on the Molecular Signatures Database version 3.0 (msigdb_v3.0.xml). Three groups were defined: tumors with detectable copy number change for a region (aberrant tumors), tumors without detectable copy number change for the same region (diploid tumors), and normal kidney. GSEA was conducted between aberrant versus diploid tumors with 10,000 permutations, minimum priori of 5 genes, and q < 0.25. Genes were assessed for significant expression changes using the Mann–Whitney U test to generate P values. A P < 0.05 and a fold-change of at least ±1.3 were selected to determine significantly expressed genes. Normalization, GSEA, and Mann–Whitney U testing were conducted using the GeneSpring GX 11 software.

**Genome-wide miRNA expression analysis**

The miRNA expression data (GSE23085) was previously described in a study from our group (19). A significance analysis of microarrays (SAM; ref. 27) was conducted and q < 0.05, fold-change of ±1.3, and numerator (r) of at least 50 were selected to determine significantly expressed miRNAs.

**Mass spectrometry protein expression analysis**

The protein expression data were previously described in a study from our group (21).

**Validation set from the cancer genome atlas**

We verified the frequency of CNAs and their impact on expression of candidate genes/miRNAs at the mRNA and protein levels in ccRCCs using the publically available "Level 3" ccRCC dataset from The Cancer Genome Atlas (TCGA), as made available through the cBio Cancer Genomics Portal (28). The TCGA dataset comprises tumors with high-resolution copy number data, with matched cases of gene (mRNA and miRNA) and protein expression. TCGA data types, platforms, and methodologies are as described previously (The Cancer Genome Atlas Research Network 2008).

**Gene-specific methylation analysis**

We also assessed methylation changes for candidate genes using matched “Level 3” TCGA ccRCC methylation data. Details of the methylation analysis can be found in the Supplementary Methods. A β-value difference of ±0.2 between tumor and normal kidney was selected as the threshold for hypermethylation (>0.2) or hypomethylation (<−0.2) as previously described (29). TCGA data types, platforms, and methodologies are as described previously (The Cancer Genome Atlas Research Network 2008).

**FISH**

FFPE matched pairs of tumor and normal tissues from 18 patients diagnosed with ccRCCs were collected, and a tissue microarray was constructed by needle dissection of 1 mm punch biopsies of tumor and normal tissues. Histopathologic diagnosis was confirmed by 2 independent pathologists. Each specimen and matched normal were represented on the tissue microarray in quadruplicate. Details of the hybridization process can be found in the Supplementary Methods.

**Quantitative real-time PCR**

Total RNA was extracted from FFPE matched pairs of tumor and normal tissues from 61 patients with ccRCCs. Total RNA extraction and quantitative real-time PCR (qRT-PCR) were carried out as previously described (30).

**Cell proliferation and migration assays**

ACHN RCC cells were transfected with synthetic miR-138 mimics. Cell transfection, proliferation, and migration assays were conducted as previously described (31). Ectopic over-expression of transfected miRNAs was verified by qRT-PCR.

**TCGA survival analysis**

We assessed select genes for CNAs and expression changes in association with survival data from TCGA as made available through the cBio Cancer Genomics Portal (28).

**Results**

**Genome-wide copy number profiling of ccRCC**

To define significant regions of copy number alteration in ccRCCs, we analyzed genome-wide copy number changes in 154 ccRCCs (see Materials and Methods; Fig. 1). We identified a total of 42 regions of either peak or broad regions of CNA in at least 5% of cases (Table 1; Fig. 2A; Supplementary Table S1). This includes 14 most significant regions of peak CNAs: 8 gains and 6 deletions (Table 1; Fig. 2A), and 28 regions (14 gains and 14 deletions) of broad aberrations that span whole chromosomes or more than half of a chromosomal arm in at least 5% of cases (Supplementary Table S1). Our results are in agreement with a number of recent studies that were conducted at high resolution (6–12), as shown in Supplementary Table S2. In addition, our analysis was able to more precisely define the boundaries of ccRCCs CNAs. Certain regions were identified in...
previous analyses such as peak deletions of 3p25.3 and 9p21.3 harboring the tumor suppressor genes VHL and CDKN2A/CDKN2B, respectively, and peak gains of 8q24.21 harboring the MYC oncogene (9).

We also identified a peak gain of 5q32, containing 5 genes including the known oncogenes CSFIR and PDGFRB, which is consistent with a recent study reporting CSFIR/PDGFRB copy number gain and overexpression of CSFIR at the transcriptomic and proteomic levels in ccRCCs (32). Moreover, 2 mutations and 1 polymorphism were previously identified in the CSFIR gene in ccRCCs, further showing this region to be a target of 5q CNAs (32). The distal region of 5q (5q34–q35.3) showed significant copy number gain as well, consistent with previous studies reporting similar boundaries of 5q distal gain in ccRCCs (8–10).

We also identified a gain of 7q22.1, harboring the MCM7 oncogene and its intronic oncogenic miRNA polyicosin, miR-106b ~ 25 in ccRCCs. Likewise, peak gain of 7p11.2 harboring 5 genes including the EGF receptor (EGFR) oncogene was infrequently observed. We also observed broad gains of chromosome 7 harboring EGFR in 17% of cases. These results are consistent with previous results of infrequent amplifications of EGFR and more frequent high polyomys and trisomy 7 in ccRCCs (33). We validated EGFR copy number gain using FISH analysis (Fig. 2C). Net copy number gains of EGFR were detected and accompanied by the concomitant gain of centromere 7, suggesting whole chromosome gain in 6 of 18 cases.

In addition, we observed a gain of 3q21.3 harboring the MCM2 oncogene. Gain of the well-characterized 12q13.2–14.1 amplicon containing the CDK4 oncogene and also the miR-26a oncogenic miRNA was observed. Gain of this region has been associated with metastatic ccRCCs (16). Rare gains of 11q13.3 harboring the CCND1 oncogene were identified as well. Amplifications of CCND1 have been observed in subsets of several cancer types (34).

Deletion of 1p36 (1p36.32–p35.3) was observed in 14% of cases. Deletion of this region is frequent in many cancer types and several candidate tumor suppressors have been identified in this region (35). We identified a deletion of the distal 2q region 2q36.1–q37.3, consistent with previous results of a high-resolution copy number analysis in ccRCCs (10). Deletions of 4q33–q35.1 and 6q23.2–q27 were defined, consistent with previous studies reporting similar boundaries of copy number loss in ccRCCs (9, 10, 12).

Broad regions of copy number gain and loss were identified as well (Supplementary Table S1), such as 1q, 5q, Chr7, Chr12, and Chr20 gain and 3p, Chr4, 6q, 8p, Chr9, and 14q deletion, consistent with previous findings (7, 8, 10, 12).

### Integrated copy number and mRNA expression analysis

 Genome-wide mRNA expression changes in ccRCCs compared with normal kidney were mapped to their respective genomic locations and correlated with CNA data (Fig. 2B). Most frequently observed was large-scale under expression of genes on 3p and overexpression of genes on 5q, which were associated with frequent deletion of 3p and gain of 5q (Supplementary Table S3). This suggests the presence of additional 3p tumor suppressor genes apart from VHL, as well as several oncogenes on 5q.

The association between gene expression and CNAs was not as clear in other regions. To better elucidate the influence of copy number changes on mRNA expression, we conducted GSEA to assess the distribution of underexpressed and overexpressed sets of genes in relation to their genomic location.
and their copy number status (see Materials and Methods). Three groups were defined: tumors with detectable copy number change for a region (aberrant tumors), tumors without detectable copy number change for the same region (diploid tumors), and normal kidney.

GSEA was run for each of the peak and broad CNAs compared with tumors without the copy number change in cases with matched gene expression data. GSEA revealed 75 gene expression sets significantly enriched in association with aberrant versus diploid tumors. Of these, 41 were associated with gains and 34 with deletions (Table 2; Supplementary Tables S4 and S5). Fifteen of these gene sets coincide with 11 significant peak regions: 6 gains and 5 deletions. This analysis shows the striking impact of CNAs on gene expression that may be overlooked when comparing the “entire” tumor population with nonmalignant tissue.

Within these gene sets, we identified numerous candidate genes with potential involvement in ccRCC pathogenesis. Statistically significant and overexpressed or underexpressed genes by at least 1.3-fold were taken into further analysis. This led to the identification of 713 concordantly gained/overexpressed and 605 concordantly deleted/underexpressed genes compared with normal kidney (Supplementary Tables S3 and S4). Similarly, 394 gained/overexpressed and 517 deleted/underexpressed genes were identified in aberrant versus diploid tumors (Supplementary Table S5).

To narrow these gene lists further, we initially focused on genes located in the peak CNAs identified by JISTIC in the GSEA gene sets. We identified 79 overexpressed genes in peak gains and 130 underexpressed genes compared with diploid tumors. Of these, 41 were associated with gains and 34 with deletions (Table 2; Supplementary Tables S4 and S5). Fifteen of these gene sets coincide with 11 significant peak regions: 6 gains and 5 deletions. This analysis shows the striking impact of CNAs on gene expression that may be overlooked when comparing the “entire” tumor population with nonmalignant tissue.

Within these gene sets, we identified numerous candidate genes with potential involvement in ccRCC pathogenesis. Statistically significant and overexpressed or underexpressed genes by at least 1.3-fold were taken into further analysis. This led to the identification of 713 concordantly gained/overexpressed and 605 concordantly deleted/underexpressed genes compared with normal kidney (Supplementary Tables S3 and S4). Similarly, 394 gained/overexpressed and 517 deleted/underexpressed genes were identified in aberrant versus diploid tumors (Supplementary Table S5).

To narrow these gene lists further, we initially focused on genes located in the peak CNAs identified by JISTIC in the GSEA gene sets. We identified 79 overexpressed genes in peak gains and 130 underexpressed genes in deletions compared with normal kidney (Supplementary Table S4). Similarly, 51 overexpressed genes in peak gains and 155 underexpressed genes in deletions compared with diploid tumors were identified (Supplementary Table S5). Among the genes in these lists, several cancer-associated genes such as CSF1R (5q32), STC2 (5q35), EGFR (7p11), MCM7 (7q22), CCND1 (11q13), and CDK4 (12q14) were significantly overexpressed in peak copy number gains. Likewise, candidate genes such as ARID1A (1p36), XRCC5/IKU80 (2q35), CDKN2AIP/CARF (4q35), SORBS2 (4q35), PERP (6q23), PLAGLI/ZAC1 (6q23), and CDKN2A

Figure 2. A, copy number profiling ccRCC. Top, heatmap and frequency plot of CNAs as visualized by the Integrative Genomics Viewer; gains are in red and losses are in blue. Bottom, JISTIC analysis of CNAs. Location of peak regions of gains and deletion (deletions, blue; gains, red). B, the correlation between mRNA expression and chromosomal aberrations in ccRCCs. Frequency of expression changes between tumor versus normal superimposed to copy number profiles (underexpression, dark blue; overexpression, dark red; deletion, light blue; gain, light red). C, copy number gains of EGFR detected by FISH. A representative example of a 2-color FISH to an FFPE tumor. The EGFR gene is labeled in red, with the centromere (CEP7) labeled in green. In most tumor cells scored for this specimen, 3 to 4 copies of EGFR/CEP7 were detected as shown in i–iii.
### Table 2. GSEA of copy number aberrant versus diploid tumors

<table>
<thead>
<tr>
<th>Cytoband&lt;sup&gt;a&lt;/sup&gt; (no. of genes/set)</th>
<th>Normalized enrichment P (q value)</th>
<th>Candidate genes&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Cytoband&lt;sup&gt;a&lt;/sup&gt; (no. of genes/set)</th>
<th>Normalized enrichment P (q value)</th>
<th>Candidate genes&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overexpression</strong></td>
<td></td>
<td></td>
<td><strong>Underexpression</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1q21 (171)</td>
<td>0.004 (0.1)</td>
<td>1.77</td>
<td>1p21 (23)</td>
<td>0.002 (0.23)</td>
<td>-1.63</td>
</tr>
<tr>
<td>1q22 (48)</td>
<td>0.007 (0.09)</td>
<td>1.77</td>
<td>1p21 (22)</td>
<td>0.002 (0.08)</td>
<td>-1.83</td>
</tr>
<tr>
<td>1q25 (46)</td>
<td>0.004 (0.1)</td>
<td>1.78</td>
<td>1p31 (56)</td>
<td>8e-4 (0.07)</td>
<td>-1.87</td>
</tr>
<tr>
<td>1q42 (59)</td>
<td>0.001 (0.007)</td>
<td>2.10</td>
<td>1p32 (49)</td>
<td>0 (0.02)</td>
<td>-1.99</td>
</tr>
<tr>
<td>1q44 (16)</td>
<td>0.001 (0.04)</td>
<td>1.90</td>
<td>1p32 (226)</td>
<td>0.03 (0.20)</td>
<td>-1.62</td>
</tr>
<tr>
<td>2q14 (25)</td>
<td>0.02 (0.21)</td>
<td>1.60</td>
<td>2q33 (52)</td>
<td>0.006 (0.04)</td>
<td>-1.84</td>
</tr>
<tr>
<td>2q21 (26)</td>
<td>0.01 (0.1)</td>
<td>1.76</td>
<td>2q35 (34)</td>
<td>0 (0.004)</td>
<td>-2.08</td>
</tr>
<tr>
<td>2q31 (48)</td>
<td>0.002 (0.08)</td>
<td>1.79</td>
<td>2q37 (67)</td>
<td>0 (0.021)</td>
<td>-1.91</td>
</tr>
<tr>
<td>3q23 (52)</td>
<td>0 (0.003)</td>
<td>2.09</td>
<td>3q24 (26)</td>
<td>0.02 (0.21)</td>
<td>1.70</td>
</tr>
<tr>
<td>3q35 (34)</td>
<td>2e-4 (0.03)</td>
<td>1.91</td>
<td>3q37 (67)</td>
<td>0 (0.03)</td>
<td>1.87</td>
</tr>
<tr>
<td>5q23 (39)</td>
<td>0.002 (0.08)</td>
<td>1.77</td>
<td>5q31 (11)</td>
<td>6e-4 (0.02)</td>
<td>1.95</td>
</tr>
<tr>
<td>5q32 (18)</td>
<td>0 (0.001)</td>
<td>1.93</td>
<td>5q34 (19)</td>
<td>0.006 (0.06)</td>
<td>2.05</td>
</tr>
<tr>
<td>5q35 (50)</td>
<td>0 (0)</td>
<td>1.80</td>
<td>5q35 (50)</td>
<td>0.006 (0.06)</td>
<td>2.05</td>
</tr>
<tr>
<td>7p11 (10)</td>
<td>0.001 (0.1)</td>
<td>1.70</td>
<td>7p12 (2)</td>
<td>0.006 (0.05)</td>
<td>1.87</td>
</tr>
<tr>
<td>7p22 (69)</td>
<td>6e-4 (0.004)</td>
<td>2.04</td>
<td>7p22 (26)</td>
<td>0.009 (0.02)</td>
<td>1.70</td>
</tr>
<tr>
<td>7q22 (36)</td>
<td>0.009 (0.1)</td>
<td>1.69</td>
<td>7q32 (26)</td>
<td>0.001 (0.1)</td>
<td>1.75</td>
</tr>
<tr>
<td>7q32 (36)</td>
<td>0.007 (0.1)</td>
<td>1.69</td>
<td>7q36 (33)</td>
<td>0.002 (0.02)</td>
<td>1.88</td>
</tr>
<tr>
<td>11q12 (74)</td>
<td>0.002 (0.008)</td>
<td>2.05</td>
<td>11q13 (174)</td>
<td>0.004 (0.07)</td>
<td>1.87</td>
</tr>
<tr>
<td>11q14 (25)</td>
<td>0.008 (0.1)</td>
<td>1.79</td>
<td>11q14 (25)</td>
<td>0.004 (0.07)</td>
<td>1.87</td>
</tr>
<tr>
<td>12p11 (23)</td>
<td>0.001 (0.1)</td>
<td>1.73</td>
<td>12p12 (28)</td>
<td>0.002 (0.02)</td>
<td>1.88</td>
</tr>
<tr>
<td>12p13 (144)</td>
<td>0 (0)</td>
<td>2.25</td>
<td>12p13 (144)</td>
<td>0 (0)</td>
<td>1.87</td>
</tr>
<tr>
<td>12p14 (25)</td>
<td>0.01 (0.15)</td>
<td>1.66</td>
<td>12p14 (25)</td>
<td>0.01 (0.15)</td>
<td>1.66</td>
</tr>
<tr>
<td>12p23 (41)</td>
<td>0.01 (0.13)</td>
<td>1.70</td>
<td>12q12 (47)</td>
<td>0.005 (0.09)</td>
<td>1.79</td>
</tr>
<tr>
<td>12q24 (130)</td>
<td>0.01 (0.13)</td>
<td>1.70</td>
<td>12q34 (27)</td>
<td>0.05 (0.19)</td>
<td>1.79</td>
</tr>
<tr>
<td>13q12 (47)</td>
<td>0.005 (0.09)</td>
<td>1.79</td>
<td>13q22 (52)</td>
<td>0.008 (0.16)</td>
<td>1.82</td>
</tr>
<tr>
<td>13q24 (52)</td>
<td>0.005 (0.09)</td>
<td>1.79</td>
<td>13q25 (34)</td>
<td>0.002 (0.16)</td>
<td>1.74</td>
</tr>
<tr>
<td>16p12 (55)</td>
<td>0.007 (0.1)</td>
<td>1.84</td>
<td>16p12 (55)</td>
<td>0.007 (0.1)</td>
<td>1.84</td>
</tr>
</tbody>
</table>

**NOTE:** Bolded regions coincide with significant focal gains and losses as detected by JISTIC. The remaining regions coincide with broad regions of gains and losses. Underlined genes were detected at the protein level by mass-spectrometry in ccRCCs; full list in Supplementary Tables S4 and S5. GSEA could not be conducted on 3p aberrant versus 3p diploid tumors; sample size limited. According to hg19. 
Select candidate genes significantly expressed in aberrant versus diploid tumors (see Materials and Methods). Full list of genes significantly expressed for each region may be found in Supplementary Tables S4 and S5.
were significantly underexpressed in peak deletions, as detailed in Fig. 3 and Discussion.

**Methylation analysis of candidate genes in regions of CNA**

Methylation data were assessed for selected genes located in regions of CNAs, revealing coordinated regulation of gene expression by both CNAs and methylation changes (Supplementary Table S6). We identified 6 genes located in peak CNAs with concordant methylation changes (2 hypomethylated in copy number gains and 4 hypermethylated in deletions). Other genes exhibiting methylation changes were located in broad regions of genomic gain and loss.

The STC2 and CCND1 genes, which are located in the 5q34-q35 and 11q13 regions of peak copy number gain, respectively, are the most overexpressed of the genes located in their region of CNA. They also showed frequent hypomethylation. It has been also shown in the literature that their expression is also
inducible by HIF1α and HIF2α, respectively (36, 37). Taken together, these data suggest the presence of multiple coordinating mechanisms (copy number gain, hypomethylation, and induction by another gene) activating the overexpression of these genes in ccRCCs. STC2 has been proposed as a prognostic marker for RCCs with increased cytoplasmic STC2 expression associated with aggressiveness and short survival time (38). Furthermore, STC2 was shown to be an oncogene that inhibits cell death in ccRCCs (6). Also, overexpression of STC2 under hypoxic conditions was previously associated with increased phosphorylation of CCND1, suggesting a biologic link between these 2 genes (37).

We confirmed frequent hypomethylation of carbonic anhydrase IX (CA9), located in 9p13, as previously observed in ccRCCs. Interestingly, we observed a marked downregulation of CA9 mRNA in 9p deleted relative to diploid tumors (fold change: −3.7). However, CA9 gene expression remained significantly overexpressed in 9p deleted tumors relative to normal kidney (fold change: 7.3), although much higher expression was observed in diploid tumors than in normal kidney (fold change: 27.0). This is of importance as higher CA9 expression has been associated with favorable prognosis and response to immunotherapy in ccRCCs (39). In addition, CA9 is HIF1α-inducible. Taken together, it appears that CA9 expression represents the net outcome of opposing (suppressing and activating) forces that can be variable among individual tumors.

We compared the expression of CA9 between hypomethylated and nonhypomethylated tumors and observed a significant increase in mRNA expression among hypomethylated tumors (Fig. 4D). We also correlated these 2 groups with TCGA survival data and observed that nonhypomethylated CA9 patients were significantly associated with worse prognosis (Fig. 4E). Furthermore, we integrated copy number loss of CA9 in conjunction with methylation status and observed copy number loss of CA9 to synergistically worsen patient survival among CA9 nonhypomethylated patients (Fig. 4F). Collectively, we observed that expression of CA9 is influenced by a combination of hypomethylation and copy number loss that significantly correlates with patient survival.

We also observed frequent hypomethylation of the nicotinamide N-methyltransferase (NNMT) gene, located in the 11q23 GSEA gene set and is broadly gained in 5% of cases. At the mRNA level, NNMT was the most highly overexpressed gene in this locus and was also upregulated at the protein level (Supplementary Tables S4 and S5). This indicates that NNMT can be activated in ccRCCs by a combination of hypomethylation and, less frequently, copy number gain. NNMT has been identified as a promising candidate diagnostic biomarker for RCCs and has been functionally observed to induce cellular invasion in ccRCCs (40).

The parathyroid hormone 1 receptor (PTH1R) gene was the most underexpressed gene on the 3p arm, consistent with the results of 2 previous datasets (6, 9). It also showed frequent
hypermethylation, suggesting the concerted role of deletion and hypermethylation in attenuating or completely silencing PTH1R expression in ccRCCs.

Likewise, a subset of tumors exhibited hypermethylation of CDKN2A and CDKN2B, located in the frequently lost 9p21, VHL (in the 3p25.3 deletion region), and SFRP1 (8p loss), consistent with previous results in ccRCCs (41, 42). In addition, in the 1p36 peak copy number deletion, hypermethylation and significant underexpression of CLCNKB were observed. CLCNKB is expressed predominantly in the kidney, and was documented to be underexpressed in kidney tumors (43), suggesting frequent hypermethylation and deletion of CLCNKB to be silencing mechanisms of its expression.

Integrated copy number and miRNA expression analysis and TCGA validation

To verify our CNA results and their impact on expression of candidate genes/miRNAs in ccRCCs, we used the cBio Cancer Genomics Portal to query the publically available TCGA ccRCC dataset (see Materials and Methods).

We previously generated genome-wide miRNA data of ccRCCs and matched normal kidney (19). We conducted a SAM and identified 145 significantly expressed miRNAs (see Materials and Methods). Integration analysis revealed 9 concordantly overexpressed miRNAs in regions of genomic gain and 6 underexpressed miRNAs in regions of deletion (Supplementary Table S7).

Interestingly, miR-26b maps to the 2q distal region that is deleted in 8% and broadly gained in 7% of cases. Further analysis of this miRNA revealed subsets of tumors either overexpressing or underexpressing miR-26b (miR-26b overexpressed in 50% and underexpressed in 20% of cases), suggesting that its expression may be copy number–dependent. We validated this relation using the TCGA dataset. We observed relative underexpression of miR-26b among copy number lost compared with cases with copy number gain status of this miRNA.

We also investigated additional miRNAs for similar relationships. miR-26a located adjacent to CDK4 was overexpressed in 25% of cases. Recent insight into the functional role of miR-26a in tumorigenesis reveals a synergistic relation between miR-26a and the CDK4 amplification in promoting aggressiveness in human cancers by cooperatively targeting multiple tumor-suppressive pathways (44). This indicates that genomic gain of the region harboring CDK4 and miR-26a in ccRCCs may contribute to this synergistic effect in tumors overexpressing miR-26a. Also, of importance to note is the location of miR-31 and miR-101-2 on the frequent copy number lost 9p arm. Expression analysis of the metastasis suppressor miR-31 revealed significant underexpression in 60%, suggesting that 9p deletions may contribute to the attenuation of miR-31 expression in a subset of ccRCC tumors.

In regard to the tumor suppressor miR-101, expression analysis revealed it to be underexpressed in 13 of 20 and overexpressed in 5 of 20 cases. miR-101 is encoded by 2 genomic loci; miR-101-2 on 9p and miR-101-1 on 1p. We observed both loci to be targeted by deletion in subsets of tumors (9p: 21% lost; 1p: 6% lost). In prostate cancer, deletion and subsequent underexpression of miR-101 led to overexpression of its oncogene target EZH2, resulting in cancer progression (45). Copy number gain of EZH2 was observed in 24% of our cases.

Our findings are in keeping with previous studies suggesting the presence of significant correlation between dysregulated miRNAs and CNAs in many cancers (46) including ccRCCs (19). The significance of miRNA dysregulation can be, in some cases, overlooked in subsets of tumors. Integration of CNAs with miRNA expression data in ccRCCs allowed us to identify new miRNAs that can be of significance in subsets of ccRCCs.

Integration with proteomics data and TCGA validation

We previously generated quantitative protein expression data by mass spectrometry comparing ccRCCs with normal kidney (21). We integrated our GSEA analysis with our mass spectrometry data and obtained 56 overexpressed genes in gains and 43 underexpressed genes in deletions compared with normal kidney that were dysregulated at the protein level in ccRCCs (Supplementary Tables S3 and S4). Similarly, 28 gained/overexpressed and 43 deleted/underexpressed genes in aberrant compared with diploid tumors were dysregulated at the protein level (Supplementary Table S5). Focusing on peak CNAs, 23 genes dysregulated at the protein level were identified; 9 associated with peak gains and 18 with deletions (Supplementary Tables S3 and S4). Most notably, of the 23 genes, PTMA, NCL, PSMD1, and XRCC5/KU80 all map to the distal 2q region that is deleted in 8% and broadly gained in 7%. Variable mRNA expression levels of these genes were observed depending on the copy number status of the region. Although this analysis provides preliminary useful information, it should be, however, interpreted with caution as the quantitative proteomic data were obtained from a separate dataset. To address this concern, we integrated our GSEA with a limited number of proteins that were analyzed on the TCGA validation set by reverse-phase protein arrays. We identified 12 proteins whose expression significantly correlates with CNAs (Supplementary Table S8). In particular, 6 of these proteins were encoded by genes located in peak CNAs (EFGR, 7p11.2; MAPK9, 5q34-q35.3; CASP9 and mTOR, 1p36.32-p35.3; ESR1, 6q23.2-q27; and XRCC5, 2q34-q37.3).

The prognostic significance of CNAs

We analyzed the correlation between CNAs with patient survival using the TCGA dataset (Fig. 4; Supplementary Table S1). Seven regions harboring key cancer-related oncogenes and tumor suppressors (Chr12, CDK4 gain; 1p, ARID1A; Chr4, SORBS2; Chr9, CDKN2A/B; 13q, RB1; 14q, NFKB1; and Chr18, DCC deletion) were significantly associated with worse prognosis, consistent with previous reports (8, 10, 13–16). We further correlated miRNA let-7i in association with available overall survival in the TCGA validation set and observed significantly worse overall survival among cases with let-7i copy number gain (Fig. 4B). Significantly higher expression of let-7i was observed in tumors with copy number gain than in diploid tumors (Fig. 4A). We further experimentally validated the correlation between let-7i expression and disease-free
survival in 61 ccRCC cases using qRT-PCR with gene-specific primers. We observed that significantly higher expression of let-7i is associated with poor prognosis (Fig. 4C).

**The effect of miR-138 on tumor proliferation and migration**

miR-138 is located on the 3p arm, a region that is the most frequently deleted in ccRCCs. In addition to VHL, studies suggest the presence of other cancer-related genes and miRNAs in this region. We identified miR-138 as a candidate tumor suppressor. We experimentally tested the effect of miR-138 on tumor characteristics using a kidney cancer cell line model. Overexpression of miR-138 in the ACHN kidney cancer cells resulted in significant reduction in the rate of cell proliferation (Fig. 5A). It also led to reduction of the migration ability of tumor cells as measured by wound-healing assay (Fig. 5B).

**Discussion**

In this study, we characterized the ccRCC genome by assessing genome-wide copy number changes and integrated gene, miRNA and protein expression, and methylation data. We identified peak aberrations harboring several oncogenes and tumor suppressors. We assessed selected genes in CNAs for methylation changes and observed concordant regulation of key oncogenes and tumor suppressors by these 2 mechanisms, such as STC2, CCND1, CA9, VHL, and CDKN2A/CDKN2B (Supplementary Table S6).

To better define the candidate genes harbored in these regions and the impact of CNAs on candidate genes, we integrated genome-wide mRNA expression data (Fig. 2B). We compared the expression of genes in tumors with a CNA to tumors without the aberration for each of the regions and identified 75 sets of genes significantly altered between these 2 groups. This led to the identification of copy number dependency of gene expression of several oncogenes and tumor suppressors in peak regions and better defined candidate genes in broad regions (Table 2; Supplementary Tables S3–S5), such as ARNT/HIF1β (1q gain), LAMP1 (Chr13 gain), CSK (Chr15 gain), TSC1 (Chr9 loss), NFKBIA (14q loss), and TP53 (17p loss).

We showed a strong association between CNAs and mRNA expression that can be overlooked because of the presence of a subset of diploid tumors for each region. These results
illustrate prestratifying patients according to their CNA pattern may help to better understand the functional impact of gene expression in specific subsets of patients.

Using this approach, we also identified new potential tumor-associated genes. For instance, the XRCC5/KU80 gene, a key mediator of double-stranded DNA break repair, showed significant overexpression in 2q copy number gain compared with diploid tumors, in addition to reduced expression among tumors harboring deletion of 2q34-g37.3 (Fig. 3A). This shows variable expression levels of KU80 among patients depending on the copy number status of the region. We validated the association between KU80 copy number status and gene expression using the TCGA dataset (Fig. 3A). We also previously showed KU80 dysregulation at the protein level by mass spectrometry in ccRCCs compared with normal kidney (21). KU80 overexpression has shown use as a biomarker to predict a higher risk of locoregional failure and death following radiotherapy in head and neck cancer, providing a rationale to stratify patients according to KU80 expression as a means to optimize course of treatment (47). This provides preliminary rationale to stratify patients according to KU80 copy number status as a means to assess treatment response. The same observation was seen for CDK4 (Fig. 3B).

Another example is the mTOR gene, which is harbored in the deleted 1p36 region. mTOR inhibitors are used for metastatic ccRCCs. Patients with mTOR deletion may not be ideal candidates for mTOR inhibitor treatment.

Furthermore, significant NFKBIA underexpression was observed in 1q4 copy number deletion compared with diploid tumors (Fig. 3C). Significant underexpression of NFKBIA/IRBa, an inhibitor of NF-κB, harbored in the frequently lost 1q4 region, was able to discriminate metastatic ccRCCs from nonmetastatic tumors in a previous study (48), suggesting deletion and subsequent underexpression of NFKBIA to be indicative of metastatic potential in ccRCCs. A recent study associated deletions and low expression of NFKBIA with resistance to treatment and worse survival in glioblastomas (49).

Our findings are in keeping with previous studies suggesting the presence of significant correlation between dysregulated miRNAs and chromosomal aberrations in many cancers (46), including ccRCCs (19). Mutations and SNPs are other layers of complexity that should be investigated and may shed more light to the complex pathogenesis of RCCs.

In conclusion, we identified new regions of peak and broad CNAs and confirmed previously reported regions of aberrations in ccRCCs that harbor potential oncogenes and tumor suppressors. Integration of multilevel molecular changes show complementarity between CNAs that may significantly influence patient survival, as is the case with C9 (Fig. 4D–F). We also identified a correlation between CNAs and gene expression (mRNA and miRNA). By stratifying patients according to their chromosomal aberration, we identified new tumor-associated genes that can be overlooked when merging the entire tumor population as one group. Finally, our results show that the same chromosomal region can harbor different classes of tumor-related molecules (such as genes and miRNAs) with a coordinated functions, as is the case with diminished tumor suppressive abilities of the frequently deleted mir-138, frequently deleted along with the tumor suppressor, VHL.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors' Contributions
Conception and design: A.H. Girgis, B. Beheshti, G.M. Yousef
Development of methodology: A.H. Girgis, V.V. Iakovlev, B. Beheshti, J. Bayani, M. Mankaruos, G.M. Yousef
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J.H. Girgis, J. Bayani, B. Khalil
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A.H. Girgis, V.V. Iakovlev, J. Bayani, J.A. Squire, M. Mankaruos, Y. Yousef, B. Khalil, G.M. Yousef
Writing, review, and/or revision of the manuscript: A.H. Girgis, V.V. Iakovlev, J. Bayani, M. Mankaruos, H.W.Z. Khella, M. Pasti, G.M. Yousef
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A.H. Girgis, A. Bui, M. Mankaruos, Y. Yousef, G.M. Yousef
Study supervision: G.M. Yousef

Acknowledgments
The authors thank Dr. Salvador Mejia-Guerrero and Rafik Matta for their assistance.

Grant Support
This work was supported by grants from the Canadian Cancer Society (CCS grant # 20185), the Ministry of Research and Innovation of the Government of Ontario, the Kidney Foundation of Canada, Canadian Institute of Health Research (CIHR grant # 2294), and the Cancer Research Society.

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.

Received February 22, 2012; revised July 28, 2012; accepted August 9, 2012; published OnlineFirst August 27, 2012.

Genomic Characterization of Clear Cell Renal Carcinoma

References


www.aacajournals.org
Cancer Res; 72(20) October 15, 2012

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2012 American Association for Cancer Research.


Multilevel Whole-Genome Analysis Reveals Candidate Biomarkers in Clear Cell Renal Cell Carcinoma

Andrew H. Girgis, Vladimir V. Iakovlev, Ben Beheshti, et al.


Updated version  Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-12-0656

Supplementary Material  Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2012/08/24/0008-5472.CAN-12-0656.DC1

Cited articles  This article cites 49 articles, 18 of which you can access for free at: http://cancerres.aacrjournals.org/content/72/20/5273.full.html#ref-list-1

Citing articles  This article has been cited by 1 HighWire-hosted articles. Access the articles at: /content/72/20/5273.full.html#related-urls

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

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

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