Integrative genomic analyses of sporadic clear cell Renal Cell Carcinoma
define disease subtypes and potential new therapeutic targets

Vijay R. Dondeti1, Bradley Wubbenhorst3, Priti Lal4, John D. Gordan1,8, Kurt D'Andrea3, Edward F. Attiyeh5,7, M. Celeste Simon1,2,6, Katherine L. Nathanson2,3

1Abramson Family Cancer Research Institute, 2Abramson Cancer Center, 3Department of Medicine, Translational Medicine and Human Genetics, 4Department of Pathology, 5Department of Pediatrics, and 6Howard Hughes Medical Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA

7Division of Oncology, Children’s Hospital of Philadelphia, PA 19104

8present address: Helen Diller Family Comprehensive Cancer Center, University of California at San Francisco, San Francisco, CA 94115, USA

Running title: Genomic analysis of ccRCC

Key words: Renal cancer, SNP arrays, DNA copy number, gene expression, DNA damage

Corresponding Author: Katherine L. Nathanson, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Room 351 BRB 2/3, 421 Curie Blvd, Philadelphia, PA 19104. Phone: 215-573-9840; Fax: 215-573-7945; E-mail: knathans@exchange.upenn.edu.

Word count: 4,957

Total Figures: 5 Total Tables: 2
Abstract

Sporadic clear cell Renal Cell Carcinoma (ccRCC), the most common type of adult kidney cancer, is often associated with genomic copy number aberrations on chromosomes 3p and 5q. Aberrations on chromosome 3p are associated with inactivation of the tumor suppressor gene von-Hippel Lindau (VHL), which activates the hypoxia inducible factors HIF1α and HIF2α. In contrast, ccRCC genes on chromosome 5q remain to be defined. In this study, we performed an integrated analysis of high-density copy number and gene expression data for 54 sporadic ccRCC tumors that identified the secreted glycoprotein STC2 (stanniocalcin 2) and the proteoglycan VCAN (versican) as potential 5q oncogenes in ccRCC. In functional assays, STC2 and VCAN each promoted tumorigenesis by inhibiting cell death. Using the same approach, we also investigated the two VHL-deficient subtypes of ccRCC, which express both HIF1α and HIF2α (H1H2) or only HIF2α (H2). This analysis revealed a distinct pattern of genomic aberrations in each group, with the H1H2 group displaying, on average, a more aberrant genome than the H2 group. Together our findings provide a significant advance in understanding ccRCC by offering a molecular definition of two subtypes with distinct characteristics as well as two potential chromosome 5q oncogenes, the overexpression of which is sufficient to promote tumorigenesis by limiting cell death.
Introduction

Approximately 210,000 individuals worldwide are diagnosed with renal cell carcinoma (RCC) each year, and RCC is responsible for more than 100,000 deaths annually (1). Based on histopathology, RCC can be classified into several types, among which clear cell Renal Cell Carcinoma (ccRCC) is the most common (2). Tumor stage and grade of ccRCC are used to stratify patients and infer prognosis (3). However, due to tumor heterogeneity within ccRCC, providing patients with reliable information about anticipated treatment response is challenging.

Some progress has been made in identifying underlying biological determinants of ccRCC. Importantly, the tumor suppressor von-Hippel Lindau (VHL) is inactivated in nearly 90% of sporadic ccRCC tumors (4, 5). pVHL, the protein encoded by VHL, is involved in multiple cellular pathways, but its best characterized function is the regulation of Hypoxia Inducible Factors HIF1α and HIF2α, key mediators of the hypoxic response (6). pVHL is the recognition component of a multiprotein complex responsible for HIF1α and HIF2α degradation under oxygen (O₂) replete conditions, and thus plays a key role in O₂ homeostasis. VHL inactivation leads to constitutive HIF activity and promotes tumor growth by enhancing angiogenesis and cell proliferation (7, 8).

While both HIF1α and HIF2α likely play significant roles in ccRCC pathogenesis, they have been shown to have some differing properties (8, 9). Both HIF1α and HIF2α promote angiogenesis, however HIF2α has been shown to be more important for tumor growth in RCC xenograft experiments (9, 10). Differences in HIFα expression have been used to classify VHL-deficient ccRCC tumors into two subtypes, with one subtype expressing both HIF1α and HIF2α (H1H2) and another expressing only HIF2α (H2) (11). Analysis of the two subtypes revealed that the H1H2 group shows increased MAPK and mTOR signaling, whereas the H2 group displays enhanced c-Myc activity. This sub-classification in part explains the heterogeneity of ccRCC (2), suggesting the possibility that the subtypes may have different clinical outcomes and need to be treated using distinct targeted therapies (11).
Cytogenetic analysis also has contributed to our understanding of ccRCC and revealed that 3p losses (60%) and 5q gains (33%) are the most prevalent genetic abnormalities in sporadic ccRCC (12, 13). Frequent VHL inactivation is in part explained by loss of 3p. However, specific targets on 5q have not yet been elucidated. Systematic sequencing of ccRCCs revealed mutations in the histone modifying genes SETD2 (SET domain containing 2), KDM5C (lysine (K)-specific demethylase 5C), and KDM6A (lysine (K)-specific demethylase 6A) and the tumor suppressor NF2 (neurofibromin 2) (14). Varela et al. identified truncating mutations in PBRM1 (polybromo 1), a SWI/SNF complex gene located on 3p (15). Of the targets identified as mutated by sequencing, only PBRM1 is involved in a large proportion (41%) of ccRCC tumors, whereas the others are present in ~3% of samples. Beroukhim et al. identified the tumor suppressors VHL, RUNX3, and CDKN2A/B as deleted and the MYC oncogene as amplified in ccRCC, however MYC amplification appears to be more important in renal cancer cell lines than in tumors (16). Most of the genes identified thus far by sequencing and copy number analysis are inactivated in ccRCC and do not affect a large percentage of cases.
Materials and Methods

Sample acquisition. Frozen tumor samples for primary analysis were obtained through the Collaborative Human Tissue Network and from the Hospital of the University of Pennsylvania. Samples were embedded in OCT and sectioned for immunohistochemistry and DNA extraction. The protocols used were approved by the University of Pennsylvania Institutional Review Board.

Immunohistochemistry. Immunostaining for HIF1α and HIF2α proteins was done as previously described in (11).

DNA extraction and copy number (CN) analysis. Total genomic DNA was extracted from tissue sections using the Promega Wizard Genomic DNA Purification Kit. Illumina HumanHap550-2v3_B (561,466 SNPs) and Human610-Quadv1.0 (620,901 SNPs) arrays were used in this study. Sample hybridization and data collection were performed at the Penn Genomics Facility according to the manufacturer’s protocol. To check for potential batch effects, we examined the hybridization controls present on all Illumina SNP arrays and box plots of signal intensities from the arrays in both batches to confirm that the data from both batches were in the same range. This analysis revealed no batch effects. The raw data were then processed using BeadStudio (Illumina) to obtain SNP-level signal intensities, which were then analyzed with Partek’s Genomic Suite to calculate SNP-level copy number. GLAD was used to segment the SNP-level CN data (17). A segment needed to contain a minimum of eight SNPs to be called a gain or loss. GLAD analysis resulted in segments with an average length of 13.5 megabases. Details about the segmentation analysis can be found in the Supplementary Methods. The segmented data was then analyzed using GISTIC (17). A total of 54 samples were analyzed; 33 new samples (GSE27852) and 21 samples (GSE13282) from a study previously published by our group (11).
Gene expression analysis. The gene expression data (GSE11904) was previously described in a study from our group (11).

Broad Data. The Broad dataset was described in (16) and downloaded from GEO (GSE14994). The segmented CN data was obtained from Tumorscape (18). Only data from sporadic ccRCC tumors was used for this analysis.

H1H2 and H2 genomic aberration analysis. The copy number (CN) data for each sample was used to determine what percent of its genome was lost (CN<1.7), normal (1.7<CN<2.3), or gained (CN>2.3). In each sample, for percentage calculations, total genome length was determined by adding up the lengths of all segments (in basepairs) provided by GLAD segmentation, and lengths of lost, normal, and gained regions were determined by adding up the lengths of all segments (in basepairs) called lost, normal, or gained, respectively. The distribution of percent genome lost, normal, or gained in H1H2 samples was compared to the corresponding distribution in H2 samples using the two-tailed t-test for statistical significance.

Cell culture, siRNAs, and transfections. The human ccRCC cell lines 786-O (obtained from ATCC) and RCC10 (kind gift from W.G. Kaelin) were used for cell culture experiments. The cell lines were verified for VHL and HIFα expression status using qPCR and Western blot analysis within the last six months. All siRNA were purchased from Ambion, except for siEZH2-1, which was purchased from Qiagen. The following siRNA were used in this study siNeg (Silencer Select Negative Control#2), siEZH2-1 (SI02665166), siEZH2-2 (s4916), siSTC2-1 (s16387), siSTC2-2 (s16388), siVCAN-1 (s229334), and siVCAN-2 (s229335). HiPerFect (Qiagen) was used to transfect cells. For all experiments, 50,000 cells were plated in 6-well plates. After 24 hours, the cells were transfected with siRNA and then grown for 3 days. All experiments were performed in standard media containing 10% serum and grown under standard conditions.
**Cell count, cell cycle, and cell death analyses.** Cells were counted three days after transfection using the Countess (Invitrogen). For cell cycle analysis, cells were pulsed with 10 μM BrdU for 15 minutes, stained with Alexa-Fluor488 conjugated anti-BrdU (Invitrogen), and then analyzed using standard protocols. For cell death analysis, cells were stained with Annexin V-FITC and PI (BioVision) and analyzed using standard protocols. For cell cycle analysis and cell death analysis, measurements were taken three days after transfection. All experiments were done in triplicate.
Results

Copy number (CN) profiling of all ccRCC. To identify aberrant genomic regions of ccRCC, we performed a genome-wide copy number analysis on 54 sporadic ccRCC tumors. Stages I, II, III, and IV comprise 46%, 15%, 30%, and 2% of the tumors in the study, respectively (Table S1). Additional clinical information for the tumors is provided in Table S1. The CN data were segmented using GLAD (17). Segments with CN>2.3 were considered “gains” and those with CN<1.7 were considered “losses”. Thresholds for gains and losses were determined by analyzing CN data of peripheral blood mononuclear cells (PBMCs). Using these thresholds, 99.5% of PBMCs’ genomes were considered “normal” (Fig. S1). Together, the gains and losses will be referred to as “aberrations”. Regions of gain on chromosomes 1 (5%), 5 (30%), 7 (17%), 11 (6%), 12 (13%), 16 (7%), 18 (6%), 19 (7%), 20 (7%), 21 (7%), and 22 (6%) and loss on chromosomes 1 (10%), 3 (80%), 6 (17%), 8 (25%), 9 (25%), 10 (11%), 13 (11%), 14 (31%), 15 (7%), and 18 (9%) were observed (numbers in parentheses indicate the percent of samples in which the aberration was present) (Fig. 1). These findings are consistent with previous reports (14, 16, 19). Although 3p losses and 5q gains were the most common alterations, considerable variation was observed (Fig. 1A). To determine the prevalence of different aberrations, a frequency histogram was plotted (Fig. 1C). Next, statistically significant regions of aberration were identified using GISTIC analysis (Fig. 1B) (17). From Fig. 1, it is evident that aberrations are prevalent throughout the entire genome in ccRCC.

Integrated copy number and gene expression analysis. The many putative target genes identified by GISTIC were too numerous to validate. Therefore, we integrated the copy number data with gene expression data to narrow the candidate gene set. Fig. 2A provides an overview of the approach to derive integrated genomic data. We applied the integrated analysis to all 54 ccRCC tumors. We further applied the method to the two subclasses - H1H2 and H2 - with the goal of identifying targets that are specific and unique to each subgroup. As the goal of the individual subtype analysis was to identify targets that are
unique to each subtype, any targets common to both subtypes were removed from the final gene list for each subtype.

**Integrated genomic analysis of all ccRCCs.** The copy number data were segmented to delineate aberrant regions. Genomic segments aberrant in at least ~10% (5 of 54) of tumor samples were selected for further analysis. Our dataset includes gene expression data for 18 tumors that underwent copy number profiling (11). The expression patterns of genes in gained segments were inspected to identify those overexpressed at the mRNA level relative to normal kidney tissue. Genes statistically significantly overexpressed by at least two-fold were taken into further analysis. This procedure led to the identification of 350 concordantly gained and overexpressed genes (Table S2). A similar comparison of lost segments with underexpressed genes resulted in the identification of 523 lost and underexpressed genes (Table S3). This set of 350 gained-overexpressed genes and 523 lost-underexpressed genes will hereafter be referred to as the “Penn Data” (Fig. 2A).

To narrow this gene list further, the Penn Data were compared to that published by Beroukhim et al. to identify targets common to both datasets (16). Beroukhim et al. investigated both sporadic and vHL disease-associated cases of ccRCC. Data from the sporadic ccRCC tumors were exclusively used for this study, as the Penn Data set only contains sporadic ccRCC tumors. The Beroukhim et al. data set contains copy number and gene expression data for 54 and 27 sporadic ccRCC tumors, respectively. Copy number and expression data were analyzed as described above for the Penn Data, leading to the identification of 453 gained-overexpressed genes (Table S4) and 586 lost-underexpressed genes (Table S5). This dataset will henceforth be referred to as the “Broad Data” (Fig. 2A). Comparison of the Penn and Broad Data sets revealed 72 gained and overexpressed target genes (Table S6) and 187 lost and underexpressed target genes (Table S7) present in both datasets (Fig. 2B). Table 1 lists the top 10 gained-overexpressed genes and top 10 lost-underexpressed genes when ordered by fold change in expression. From Table 1, it is evident that all the targets are aberrant in more than 10% of ccRCCs.
Copy number analysis of ccRCCs based on HIF expression. We have previously shown that VHL-deficient ccRCC tumors can be grouped into two subtypes - H1H2 and H2 - based on HIFα expression (11). Immunostaining was used to classify the 54 ccRCC tumors in this study into 29 H1H2, 19 H2, and 6 VHL wildtype tumors (Fig. S2). The VHL wildtype tumors were not analyzed further due to their low number. Using the available copy number data, we profiled the gains and losses of the two subtypes. To determine the overall prevalence of aberrations, a frequency histogram was plotted for each subtype (Fig. 3A). Differences between H1H2 and H2 tumors were observed. Gains are found over the entirety of chromosome 5 in H1H2 tumors, whereas in H2 tumors they localized to 5q. Only H1H2 tumors appear to have gains on chromosomes 16 (9%), 19 (7%), 20 (5%), and 22 (3%). The differences between the two subtypes are even more marked when losses are cataloged. Losses on chromosome 6q and 8p are more frequent in H1H2 tumors than in H2 tumors, at 30% and 40%, and 3% and 5%, respectively. GISTIC analysis was performed to identify statistically significant aberrations within each group (Fig. 3B and 3C). The results from GISTIC analysis were consistent with the findings from the frequency histograms and indicate that both groups differ in copy number aberration pattern, supporting that each is distinct.

Integrated genomic analysis of H1H2 and H2 tumors. To identify candidate genes affecting cancer in each subtype, copy number profiling was combined with gene expression data for each subtype. The integrated genomic analysis was performed as described above, but was carried out independently for each subtype. The goal of performing the integrated analysis on the two subtypes individually was to identify targets that are unique and specific to each subtype. Hence, any targets common to both subgroups were removed. Removing targets common to both subgroups helps us define targets that may specifically play a role in one subtype but not the other. This analysis uncovered 48 gained-overexpressed genes (Table S8) and 106 lost-underexpressed genes (Table S9) that are restricted to the H1H2 subtype and 121 gained-overexpressed genes (Table S10) and 79 lost-underexpressed genes (Table S11) that are restricted to the H2 subtype. Table 2 lists the top five gained-overexpressed genes and the top five lost-underexpressed genes from each subtype when sorted by fold change in expression. This analysis of
H1H2 and H2 tumors suggests that there are factors that influence tumorigenesis unique and specific to each subtype.

**Comparison of level of genomic aberration in H1H2 and H2 tumors.** We previously noted that a key difference between the two subtypes was the overexpression of DNA damage response genes, particularly those involved in double strand break repair, in H2 samples relative to H1H2 samples (11). This finding led us to hypothesize that H2 tumors would exhibit fewer genomic aberrations compared to H1H2 tumors, which was supported preliminarily by copy number profiling of 10 H1H2 and 11 H2 tumor samples (11). Herein, we use a larger sample set and more detailed copy number (CN) distribution analysis in order to confirm our initial findings. Using the segmented CN data of the H1H2 and H2 tumors, the average CN distribution for each subtype was plotted (Fig. 4). The CN distribution depicts the average percent of the genome present in a given copy number bin (of 0.1 from 0 to 4) in each subtype. From Fig. 4, it is apparent that in H2 tumors a greater percentage of the genome can be described as normal when compared to H1H2 (1.7<CN<2.3, P=0.009). Conversely, in H1H2 tumors a greater percentage of the genome can be regarded as lost (CN<1.7, P=0.005). Also, on average a greater percentage of the genome of H1H2 tumors was gained (CN>2.3) compared to H2 tumors, and this finding showed a trend towards significance (P=0.064). This detailed analysis further establishes that H1H2 and H2 tumors are indeed two distinct subgroups of ccRCC. Thus the two subtypes have different gene expression profiles, copy number profiles, and genomic aberration patterns.

**Validation of integrated genomic analysis.** The results above indicate that the integrated genomic approach can be used to identify targets that are important for ccRCC as a whole and for H1H2 and H2 subtypes individually. Next, we wanted to validate the integrated genomic approach by functionally validating selected putative targets in cell culture, and focused on those genes identified through our analysis of the whole set of ccRCCs. Following a review of the gained-overexpressed targets, \( EZH2 \) (7q36.1), \( STC2 \) (5q35.2), and \( VCAN \) (5q14.3) were chosen for further investigation. \( EZH2 \) was selected
since it has been shown to be an unfavorable prognostic marker in ccRCC (20) and inhibits apoptosis in ccRCC cells (21). STC2 was chosen as it has been found to be a negative prognostic marker in ccRCC (22), but its functional role in ccRCC has not yet been defined. Finally, VCAN was chosen because, although it has not yet been linked to ccRCC, it has been demonstrated to be upregulated in ovarian cancer (23) and promote cell proliferation and inhibit apoptosis in cell culture assays (24). STC2 is at the distal region of the frequently gained 5q region, whereas VCAN is at the proximal region. We chose STC2 and VCAN so that we could evaluate two different and distant regions of the 5q gain.

Validation of potential targets: EZH2, STC2, and VCAN. To evaluate the potential roles played by EZH2, STC2, and VCAN in ccRCC, siRNA experiments were performed using 786-O cells (expressing only HIF2α, “H2”) and RCC10 cells (expressing HIF1α and HIF2α, “H1H2”), human ccRCC cell lines. Two independent siRNAs each were used to silence EZH2, STC2 and VCAN. Real-time quantitative Reverse Transcription PCR (qRT-PCR) was used to monitor knockdown efficiency (Fig. 5A). Western blot analysis was used to confirm decrease in protein levels after silencing the targets (Fig. S3C). First, we examined whether inhibiting these targets affects cell number. Suppressing EZH2, STC2, and VCAN significantly reduced cell numbers relative to the negative control siRNA in 786-O and RCC10 cells (Figs. 5B and S3A). To determine whether the targets affected cell proliferation or cell death, cell cycle studies and Annexin V staining analysis were undertaken (Figs. 5E, 5F, and S4). Interestingly, cell cycle analysis failed to reveal any marked differences between cells treated with negative control and target siRNA (Fig. 5C). However, both 786-O and RCC10 cells treated with target siRNA showed a significant increase in cell death when compared to those exposed to negative control siRNA (Figs. 5D and S3B). Silencing targets of interest increased cell death by 4-15%. These results indicate that EZH2, STC2, and VCAN primarily promote cell growth by inhibiting cell death. These experimental findings validate the integrated genomic approach used to identify the candidate genes.

We also tested whether simultaneous inhibition of STC2 and VCAN would have an additive effect on cell numbers. Upon simultaneously silencing STC2 and VCAN, we achieved >70% knock-down.
efficiency for both targets (Fig. S5A), but did not observe any additive or synergistic effect on cell viability (Fig. S5B and S5C). We also employed TUNEL staining to determine what effect gain of \textit{EZH2}, \textit{STC2}, or \textit{VCAN} may have on apoptosis in tumors. We performed TUNEL staining on six tumors without changes in the three targets and 12 tumors with a gain in one or more of the selected loci, comparing the percentage of nuclei that were TUNEL positive (Fig. S6). We did not see a statistically significant difference in TUNEL positivity between tumors with a gain of at least one of the targets and those without. Tumors lacking gains in the \textit{EZH2}, \textit{STC2}, or \textit{VCAN} loci may therefore possess other factors inhibiting apoptosis.
Discussion

The goal of this study was to use an integrated genome-wide approach to identify genes which play an important role in sporadic ccRCC and two individual ccRCC subtypes H1H2 and H2. Copy number analysis was integrated with gene expression data for 54 sporadic ccRCCs to find genes that were either concordantly gained and overexpressed, or lost and underexpressed. The same analysis was also performed on a publicly available dataset (Broad Data) (16), and the results compared to find overlapping targets. This process led to the identification of 72 gained-overexpressed and 187 lost-underexpressed genes.

We also used the integrated genomic approach to study the differences between the H1H2 and H2 sub-sets. Although H1H2 and H2 tumors share some overlapping copy number and expression changes, the integrated genomic analysis revealed that there are genes involved in tumorigenesis that are unique and specific to each subtype. Using copy number analysis, we found that the genome of the H1H2 group is on average more aberrant than the H2 group. This difference may be because H2 tumors express DNA damage response genes, particularly those involved in double strand break repair, at a higher level than the H1H2 group (11). The genomic differences in H1H2 and H2 tumors may have clinical implications. GISTIC analysis revealed that copy number losses in 9p and 14q are more significant in the H2 group compared to the H1H2 group. However, at the level of gene expression we only see significant changes in the genes from 14q in the H2 subtype. Intriguingly, losses of 14q have been independently associated with a decrease in disease-specific survival in ccRCC (12). The data presented herein support the postulate that ccRCC can be subtyped based on HIFα expression; the survival data from Klatte et al. indicate that the H2 subtype may be potentially linked to clinical outcome (12). Together, these findings potentially suggest that different therapeutic regimens may need to be employed to treat patients with H1H2 and H2 tumors.

In order to validate the integrated genomic approach, we chose to study three of the identified targets using cell culture studies. Genome-wide studies of ccRCC to-date have primarily revealed genes
which are inactivated during tumorigenesis. Thus, we focused on the concordantly gained and overexpressed genes. Three gained-overexpressed genes, \textit{EZH2} (7q36.1), \textit{STC2} (5q35.2), and \textit{VCAN} (5q14.3) were chosen for study in cell culture experiments. \textit{EZH2} is thought to promote cell proliferation and inhibit cell differentiation by silencing tumor suppressors (25, 26), and known to be activated in breast (27) and prostate cancers (28). It also has been implicated in ccRCC (20, 21). It is interesting that while \textit{EZH2}, a histone methyltransferase, is gained, \textit{KDM6A}, a histone demethylase, is mutated in ccRCC (14). \textit{EZH2}, a member of the PcG family, is believed to play a role in cancer by silencing tumor suppressors, such as \textit{ARF} (26, 29). KDM6A has been shown to demethylate many RB binding proteins leading to their activation and subsequent cell cycle arrest (30). Thus both \textit{KDM6A} mutation and \textit{EZH2} gain will lead to increased cell cycle progression. \textit{STC2} has been shown to be overexpressed in prostate (31), breast (32), and colorectal cancers (33), and has been linked to ccRCC (22), but its functional role has not yet been determined. \textit{STC2} is upregulated under hypoxia, and thought to help cells adapt to the stress of the tumor microenvironment (33). \textit{VCAN} is involved in cell adhesion, proliferation, migration, and angiogenesis, and thought to promote cell proliferation by increasing the propagation of signals from mitogens such as \textit{platelet-derived growth factor (PDGF)} and \textit{transforming growth factor beta 1 (TGF-β1)} (34). \textit{VCAN} has been linked to prostate, breast, and ovarian cancers (23, 34), but not yet to ccRCC. Using siRNA experiments, we demonstrated that \textit{EZH2}, \textit{STC2}, and \textit{VCAN} promote tumor growth by inhibiting cell death in ccRCC cells. These results strongly suggest that \textit{EZH2}, \textit{STC2}, and \textit{VCAN} play roles in ccRCC and validate the integrated approach employed to identify them.

While the experiments described here examine \textit{EZH2}, \textit{STC2}, and \textit{VCAN} individually, it is evident from the copy number data that the genes are gained simultaneously. In tumors that show a gain of \textit{STC2}, there is a 29% and 48% chance that \textit{EZH2} or \textit{VCAN} will also be gained, respectively. In ~4% of the tumors all three targets are gained. In order to test whether \textit{STC2} and \textit{VCAN} have an additive effect on cell numbers and cell death, we simultaneously silenced them. We did not detect any additive or synergistic effects. More studies will be required to better understand the detailed functions of \textit{EZH2}, \textit{STC2}, and \textit{VCAN} and whether they interact \textit{in vivo}.
EZH2, STC2, and VCAN are commonly gained in both the H1H2 and H2 subtypes of ccRCC. EZH2, STC2 and VCAN are gained in 7%, 31%, and 24% of H1H2 samples respectively, and in 11%, 32%, and 11% of H2 samples respectively. In addition to these common targets, the individual subtype analysis has revealed that there are targets that are unique to each subtype. EZH2, STC2, and VCAN may be responsible for the earlier steps in tumorigenesis in both subtypes, whereas the targets that are specific to each subtype may be more important in the later steps of tumorigenesis. More work is needed to understand the different roles played by the common targets and roles played by the targets specific to each subgroup. It is also important to note that most (46%) of the tumors in this study are Stage I tumors, and that it is possible that alternative targets may be important during different tumor stages.

In summary, we are the first to identify and functionally validate two potentially important targets on 5q (STC2 and VCAN), a region gained in more than 30% of ccRCC samples. We also have further established that ccRCC can be classified into subtypes based on HIFα expression with each group having its own specific pattern of copy number alterations.
Acknowledgments

We thank John Maris, Brad Johnson, and Frank Lee for useful discussions; Penn Genomics Facility for sample hybridization and data collection; Hakon Hakonarson for providing CN data for PBMCs. Tissue samples were provided by the Cooperative Human Tissue Network, which is funded by the National Cancer Institute. Other investigators may have received samples from these same tissues.

Grant Support

This work was supported by NIH Program Project Grant CA104838 (to MCS), the Howard Hughes Medical Institute (to MCS), NIH R01 Grant CA135509 (to KLN), and the Abramson Cancer Center (ACC). This project is funded, in part, under a grant with the Pennsylvania Department of Health (to KLN). The Department specifically disclaims responsibility for any analyses, interpretations, or conclusions. MCS is an Investigator of the Howard Hughes Medical Institute. Services provided in the Penn Genomics Facility are supported by the ACC core grant 5P30CA016520.
References


Table 1. Genes determined to be either gained and overexpressed or lost and underexpressed in both the Penn Data and Broad Data sets. Frequency indicates how often the gene is aberrant in the Penn Data set.

<table>
<thead>
<tr>
<th>Gained and Overexpressed</th>
<th>Gene</th>
<th>Cytoband</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPCAT1</td>
<td>5p15.33</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>VCAN</td>
<td>5q14.3</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>HAPLN1</td>
<td>5q14.3</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>LOX</td>
<td>5q23.1</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>TGFB1</td>
<td>5q31.1</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>FABP6</td>
<td>5q33.3</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>PTG1</td>
<td>5q33.3</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>STC2</td>
<td>5q35.2</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>NPTX2</td>
<td>7q22.1</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>EZH2</td>
<td>7q36.1</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lost and Underexpressed</th>
<th>Gene</th>
<th>Cytoband</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTH1R</td>
<td>3p21.31</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>LPA</td>
<td>6q25.3</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>PLG</td>
<td>6q26</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>LRRC19</td>
<td>9p21.2</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>TYRP1</td>
<td>9q23</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>TRPM3</td>
<td>9q21.11</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>FBP1</td>
<td>9q22.32</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>ALDOB</td>
<td>9q31.1</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>SLC7A8</td>
<td>14q11.2</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>SERPINA5</td>
<td>14q32.13</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Genes determined to be either gained and overexpressed or lost and underexpressed uniquely in H1H2 or H2 subtypes. Frequency indicates how often the gene is aberrant.

<table>
<thead>
<tr>
<th>H1H2</th>
<th>Gained and Overexpressed</th>
<th>Lost and Underexpressed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gene</td>
<td>Cytoband</td>
</tr>
<tr>
<td></td>
<td>UBE2I</td>
<td>16p13.3</td>
</tr>
<tr>
<td></td>
<td>C16orf42</td>
<td>16p13.3</td>
</tr>
<tr>
<td></td>
<td>TPM4</td>
<td>19p13.12</td>
</tr>
<tr>
<td></td>
<td>AP3D1</td>
<td>19p13.3</td>
</tr>
<tr>
<td></td>
<td>GADD45B</td>
<td>19p13.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H2</th>
<th>Gained and Overexpressed</th>
<th>Lost and Underexpressed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gene</td>
<td>Cytoband</td>
</tr>
<tr>
<td></td>
<td>ELL2</td>
<td>5q15</td>
</tr>
<tr>
<td></td>
<td>C5orf13</td>
<td>5q22.1</td>
</tr>
<tr>
<td></td>
<td>EPB41L4A</td>
<td>5q22.22</td>
</tr>
<tr>
<td></td>
<td>RAPGEF6</td>
<td>5q31.1</td>
</tr>
<tr>
<td></td>
<td>HAVCR1</td>
<td>5q33.33</td>
</tr>
</tbody>
</table>
Figure Legends

Fig. 1. Copy number profiling of all ccRCCs. (A) Heatmap of copy number alterations of 54 sporadic tumors. (B) GISTIC analysis of copy number changes in tumors. (C) A frequency histogram showing the percentage of samples that are altered at each genomic segment. Gains are in red and losses are in blue.

Fig. 2. Summary of the integrated genomic analysis. (A) Flowchart illustrating the steps in the integrated genomic analysis. (B) Venn diagrams showing the number of targets gained and overexpressed or lost and underexpressed in the Penn Data and Broad Data sets.

Fig. 3. Copy number profiling of H1H2 and H2 subtypes. (A) Frequency histograms showing the percentage of samples that are altered at each genomic segment. (B) GISTIC analysis of the gained regions in H1H2 and H2 tumors. (C) GISTIC analysis of the lost regions in H1H2 and H2 tumors. Gains are in red and losses are in blue.

Fig. 4. Copy number (CN) distribution in H1H2 and H2 subtypes. In H2 tumors, a greater percentage of the genome is normal (1.7<CN<2.3, P=0.009) compared to H1H2 tumors, and conversely, in H1H2 tumors a greater percentage of the genome is gained (CN>2.3, P=0.064) or lost (CN<1.7, P=0.005) relative to H2 tumors.

Fig. 5. Validation of EZH2, STC2, and VCAN in 786-O cells. (A) Verification of knockdown by qRT-PCR. (B) Reduced cell numbers observed after silencing targets. (C) No significant change in cell proliferation seen after silencing targets. (D) Increased cell death observed after silencing the targets. (E) Representative plots of BrdU incorporation studies. (F) Representative plots of FITC-Annexin V and PI staining analysis. * represents P<0.05.
Figure 1

Dondeti et al. (2011)
Identify Gained and Lost Segments (in at least 10% of samples)

Identify Overexpressed and Underexpressed Genes (at least a two-fold change)

Identify Genes that show concordant changes in Copy Number and Gene Expression

“Penn Data”

“Broad Data”

Gained and Overexpressed

Lost and Underexpressed

Penn Data

Broad Data

278

72

381

399

B

Gained and Overexpressed

Lost and Underexpressed

Penn Data

Broad Data

336

187

399

Figure 2 Dondeti et al. (2011)
Figure 4

Dondeti et al. (2011)
Figure 5

(A) Relative mRNA levels

(B) Relative cell numbers

(C) % BrdU incorporation

(D) % Cell death

(E) DNA BrdU

(F) FITC-Annexin V

Dondeti et al. (2011)
Integrative genomic analyses of sporadic clear cell Renal Cell Carcinoma define disease subtypes and potential new therapeutic targets

Vijay R Dondeti, Bradley Wubbenhorst, Priti Lal, et al.

Cancer Res Published OnlineFirst November 17, 2011.

Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-11-1698

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2011/11/17/0008-5472.CAN-11-1698.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

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.