CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Yale University

Yale University
EliScholar – A Digital Platform for Scholarly Publishing at Yale

Yale Medicine Thesis Digital Library | School of Medicine

January 2016

Genomic Characterization Of Sarcomatoid Transformation In Clear Cell Renal Cell Carcinoma

Xiao Bi
Yale University, xiao.bi@yale.edu

Follow this and additional works at: http://elischolar.library.yale.edu/ymtdl

Recommended Citation
Bi, Xiao, "Genomic Characterization Of Sarcomatoid Transformation In Clear Cell Renal Cell Carcinoma" (2016). *Yale Medicine Thesis Digital Library*. 2039.
http://elischolar.library.yale.edu/ymtdl/2039

This Open Access Thesis is brought to you for free and open access by the School of Medicine at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Yale Medicine Thesis Digital Library by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

Genomic Characterization of Sarcomatoid Transformation in Clear Cell Renal Cell Carcinoma

A Thesis Submitted to the

Yale University School of Medicine

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Medicine

By

Xiao Bi

2016

Abstract

GENOMIC CHARACTERIZATION OF SARCOMATOID TRANSFORMATION IN CLEAR CELL RENAL CELL CARCINOMA. Mark Bi[a], Siming Zhao[a], Jonathan W. Said[e], Maria J. Merino[g], Adebowale J. Adeniran[d], Zuoquan Xie[b], Cayce B. Nawaf[b], Jaehyuk Choi[h], Arie S. Belldegrun[f], Allan J. Pantuck[f], Harriet M. Kluger[c], Kaya Bilguvar [a], Richard P. Lifton[a], Brian Shuch[b]. [a]Department of Genetics and Howard Hughes Medical Institute, [b]Department of Urology, [c]Division of Oncology, Department of Medicine, [d]Department of Pathology, [h]Department of Dermatology, Yale School of Medicine, New Haven, CT. [e]Department of Pathology, [f]Department of Urology, UCLA School of Medicine, Los Angeles, CA. [g]Translational Surgical Pathology Division, National Cancer Institute, Bethesda, MD

The presence of sarcomatoid features in renal cell carcinoma (ccRCC) confers a poor prognosis. The mechanisms that account for these sarcomatoid features are unknown. We performed whole exome sequencing of matched normal-carcinomatous-sarcomatoid specimens from 21 subjects. Two tumors had hypermutation and a mutational signature consistent with mismatch repair deficiency. In the remaining 19 tumors, sarcomatoid and carcinomatous elements shared a mean of 45/108 (41.7%) somatic single nucleotide variants (SSNVs). Sarcomatoid elements had a higher overall SSNV burden (mean 90 vs 63 SSNVs, $p=4.0x10^{-4}$), increased frequency of non-synonymous SSNVs in Pan-Cancer genes (mean 1.4 vs 0.26, p=0.002), and increased frequency of loss of

heterozygosity across the genome (median 913 vs 460 Mb in LOH, $p < 0.05$), with significant recurrent segments of LOH on chromosomes 1p, 9, 10, 14, 17p, 18, and 22. The most frequent somatic mutations shared by carcinomatous and sarcomatoid elements were in known ccRCC genes (*VHL*, *PBRM1*, *SETD2, PTEN)*. Most interestingly, sarcomatoid elements acquired new bi-allelic *TP53* mutations in 32% of tumors ($p=5.47x10^{-17}$); *TP53* mutations were absent in carcinomatous elements in non-hypermutated tumors and rare in previously studied ccRCCs. Mutations in known cancer drivers *ARID1A* and *BAP1* were significantly mutated in sarcomatoid elements, and were mutually exclusive with *TP53* and each other. Additionally, LOH on chromosome 9 was found in all TP53-mutant tumors. These findings demonstrate that sarcomatoid elements arise from dedifferentiation of carcinomatous ccRCCs and implicate specific genes in this process. These findings have implications for the treatment of patients with these poor-prognosis cancers.

Acknowledgements:

I would like to sincerely thank my faculty sponsors, Richard Lifton and Brian Shuch, for their mentorship and for the opportunity to work on such exciting projects. I would also like to express my gratitude to Siming Zhao, another graduate student in the Lifton lab, for her close guidance and teaching during my time there, as well as to the other members of the Lifton lab, all of whom contributed to my education and growth.

I received financial support from the NIH CTSA TL1 Medical Student Research Fellowship, without which this would not be possible. I would also like to acknowledge the Yale Office of Student Research for their work in facilitating medical student research.

Lastly, I would like to thank the other co-authors on the study not named thus far for their valuable contributions to the following body of work, as well as to the staff of the Yale Center for Genome analysis for their technical excellence in sample preparation, capture and DNA sequencing.

Table of Contents

Introduction

Sarcomatoid transformation is common in various epithelial malignancies, featuring further loss of differentiation and acquisition of characteristics typical of a sarcoma. In renal cell carcinoma (RCC), sarcomatoid features are observed in 5% of tumors. However, among individuals with stage IV disease, it occurs in 15%.(1, 2) While once believed to represent a distinct subtype of RCC, it is now considered a specific histologic feature.(3) While sarcomatoid features are found in all forms of kidney cancer, >65% of cases are found with clear cell RCC (ccRCC).(4, 5) When sarcomatoid features are present, renal tumors generally are large (median size 10 cm), invasive (20%), and/or metastatic (50%) at presentation.(4, 5) While all such tumors are considered to be Fuhrman grade IV,(6) their prognosis is significantly worse when compared to other high-grade tumors. (7) These tumors, when metastatic, have amongst the poorest survival of all genitourinary malignancies, with a median survival of only six months.(1, 2, 4) Even with resected localized disease, nearly 75% recur and have a median survival of <2 years.(4, 8) The response to systemic therapy is poor, with rare durable responses occurring with any therapeutic strategy.(9-12)

Understanding of the genetic events associated with sarcomatoid kidney tumors is currently limited.(9) Various theories have been proposed regarding the origins of sarcomatoid features in renal tumors. Given that they virtually always occur in conjunction with typical epithelial RCC elements, the terminology of a “mixed malignancy” appeared a half century ago.(13) Proposals have included

independent occurrences of tumor types in close proximity, as has been observed in various genitourinary malignancies (14), and the influence of tumor microenvironment.(15) The current prevailing theory is that sarcomatoid features represents a sub-clonal de-differentiation or transformation from an incident carcinomatous component.(16) However, the current theory is based on limited evidence. Evidence of common cell of origin is suggested by the shared patterns of X-chromosome inactivation.(17) While there is limited evidence that the epithelial component transforms into the sarcomatoid element, groups have considered sarcomatoid features to result from a final common de-differentiation pathway in RCC.(16) This is based on the sarcomatoid component more frequently metastasizing, possessing higher tumor grade, increased proliferative index, and frequent reduced expression of epithelial adhesion molecules such as E-cadherin.(18-20) As some tumors demonstrate increased expression of N-Cadherin, it has been suggested that epithelial-mesenchymal transformation (EMT) may be involved in the development of sarcomatoid elements,(21). However, what could be driving this process has thus far escaped elucidation.

The use of next generation sequencing, including whole exome sequencing and whole genome sequencing, has brought dramatic advancements in the understanding of the basic biology of a multitude of cancer types, among them clear cell renal cell carcinomas. A number of commonly mutated genes have been identified in ccRCCs, chief among them the von Hippel-Lindau tumor suppressor (*VHL*), which promotes angiogenesis under conditions of hypoxia(22). Several genes involved in chromatin modification are also commonly implicated, including

polybromo 1 (*PBRM1*) and AT-rich interactive domain-containing protein 1A (*ARID1A*), both components of the Switch/Sucrose NonFermentable (SWI/SNF) chromatin remodeling complex, SET domain containing protein 2 (*SETD2*), and lysine-specific demethylase 5C (*KDM5C*)(22). One study found genes involved in the ubiquitin-mediated proteolysis pathway (UMPP) to be the most frequently mutated, most notably the tumor suppressor BRCA1 associated protein-1 (*BAP1*)(23). In addition to gene mutations, chromosomal copy number changes are also common in ccRCCs, with the most common event being loss of chromosome 3p. Notably, chromosome 3p contains the genes *VHL, PBRM1,* and *BAP1.*(24)

Despite the success of next generation sequencing in furthering our understanding of ccRCCs, to date no comprehensive genomic studies have been conducted on the process of sarcomatoid transformation in this cancer type. We therefore sought to study this deadly event using next generation sequencing techniques.

Statement of Purpose

In an effort to elucidate the molecular basis for sarcomatoid transformation in ccRCC, we performed whole exome sequencing of distinct regions of clear cell and sarcomatoid morphology from the same tumors in a cohort of patients with ccRCC with regions of sarcomatoid transformation, using adjacent normal kidney tissue as matched control. Through comparison of normal, clear cell, and sarcomatoid exome sequencing results, we aim to address the following:

1. Define the evolutionary relationship between clear cell and sarcomatoid elements in individual tumors. Comparison of somatic mutations in clear cell and sarcomatoid components will reveal if they arise as separate independent entities or if one component evolves as a sub-clonal outgrowth of another component.
2. Characterize the landscape of mutational burden in sarcomatoid tumor components, particularly in comparison to clear cell elements, including frequency and type of single nucleotide variants as well as regions of chromosomal deletion leading to loss of heterogeneity.
3. Characterize specific genes implicated in the process of sarcomatoid transformation.

Materials and Methods

Patients and Specimen Acquisition. From 1989 to 2010, all patients undergoing nephrectomy for presumed renal cancer at the University of California, Los Angeles (UCLA) had clinical information entered into an approved database. ccRCC tumors featuring sarcomatoid transformation were reviewed by a genitourinary (GU) pathologist (JWS). The clinical data, tumor characteristics, and survival have previously been described.(1, 4, 20) A second GU pathologist (MJM) reviewed representative slides in order to confirm the presence of distinct morphologic regions that represented 1) ccRCC and 2) a definitive region with sarcomatoid transformation.

All living patients studied provided written informed consent for participation in this research. A waiver of consent was approved to study anonymized samples from deceased patients. The research protocol was approved by the UCLA and Yale Human Investigation Committees. All experiments were conducted according to the principles expressed in the Declaration of Helsinki.

DNA Extraction, Exome Capture, and Exome Sequencing. Twenty-nine tumors had available formalin-fixed, paraffin-embedded blocks with adjacent normal kidney for genomic control. The blocks were reviewed by a third independent GU pathologist (AA), who confirmed the histology and identified distinct regions of normal kidney, clear cell, and sarcomatoid histology. From each of these regions, 1 mm punches were obtained and DNA extracted using a previously described protocol.(25) Exome capture was performed using Nimblegen 2.1M Human

Exome Array followed by 74 base paired-end DNA sequencing on the Illumina HiSeq instrument. Tumor regions were sequenced to greater depth of coverage than normal tissue in order to account for admixture of tumor and normal cells. High quality sequences were obtained for twenty-one matched sets of normal, carcinomatous and sarcomatoid elements.

Sequence Analysis and Comparisons. Sequences were aligned to the hg19 reference genome using the Burrows-Wheeler Aligner (BWA)-MEM algorithm(26). From this, somatic single nucleotide variant (SNV) calling was performed using Mutect(27) and indel calling was performed using Indelocator (https://www.broadinstitute.org/cancer/cga/indelocator). Additional somatic mutation calls were acquired using a previously reported pipeline, based on a Fisher's Exact Test of reference and non-reference read counts in tumor versus normal specimens (28). Variant calls with less than a total of 8 independent reads in any of the three sequenced samples from each patient were discarded. Variants previously identified as germline variants in 1000 Genomes Project,(29) National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project, and the Yale University exome database were excluded. All somatic mutation calls were manually verified by visual inspection.

Somatic mutations identified by Mutect in one component (e.g. the sarcomatoid but not carcinomatous elements) were called specific for that component when the variant was not called in the other component, had $p \leq 0.05$ for two-tailed Fisher's exact test of the difference in the distribution of reference

and non-reference reads between the sarcomatoid and carcinomatous components, and MAF < 10% or non-reference read count <3 in the component lacking the variant. Empirically, component-specific variants had median counts of 14 independent variant reads in the component in which they were called and 0 variant reads in the component in which they were absent. All other somatic mutations were called in both components or were not excluded and were classified as shared. Phylogenetic trees were constructed for each sample based on the component distributions of somatic mutations.

Gene burden analyses were conducted to identify genes with a higher than expected number of somatic mutations predicted to affect protein function. Somatic mutations predicted to affect protein function were defined as nonsense, splice site, and small insertion/deletion (indel) variants, as well as missense variants in evolutionarily conserved amino acid positions and missense variants predicted as damaging by the PolyPhen-2 algorithm(30). Shared, carcinomatous component specific, and sarcomatoid component specific mutations were analyzed separately. We adjusted for gene size and expression in kidney epithelium to account for the effects of transcription-coupled DNA repair(31). A significance threshold of $p=0.05$ was used for known Pan-Cancer genes and $p=2.5x10^{-6}$ for all other genes to account for multiple testing of 25,000 genes across the genome.

Chromosomal segments with loss of heterozygosity (LOH) were identified from departure of the minor allele frequencies of heterozygous SNPs in tumor samples from the frequencies seen in matched normal samples(28).

Chromosomal arms with elevated rates of LOH specific to sarcomatoid components were identified using a binomial distribution with FDR < 0.25. Tumor purities were estimated using the difference in allele frequencies between tumor and normal components in regions of LOH.

Results in our cohort were compared to sequencing results seen in 424 ccRCCs in The Cancer Genome Atlas (TCGA) ccRCC study (22). Local pathology reports of TCGA samples were reviewed to identify samples with sarcomatoid features. Data from cBioPortal was used to determine the frequency of specific mutations, overall mutational burden, and concomitant LOH in these samples.(32)

Results

Exome sequencing of carcinomatous and sarcomatoid elements of ccRCCs. The clinical and pathologic features of the twenty-one patients with ccRCC with sarcomatoid transformation are shown in Table 1. Cancer-specific survival was poor, with a 1 and 2-year survival of 38 and 30%, respectively (Figure 1). Similar to other cohorts, tumors were large (median 10 cm), were frequently associated with metastases (66.6%), and frequently showed local invasion (80.9% T3/T4; Table 1).

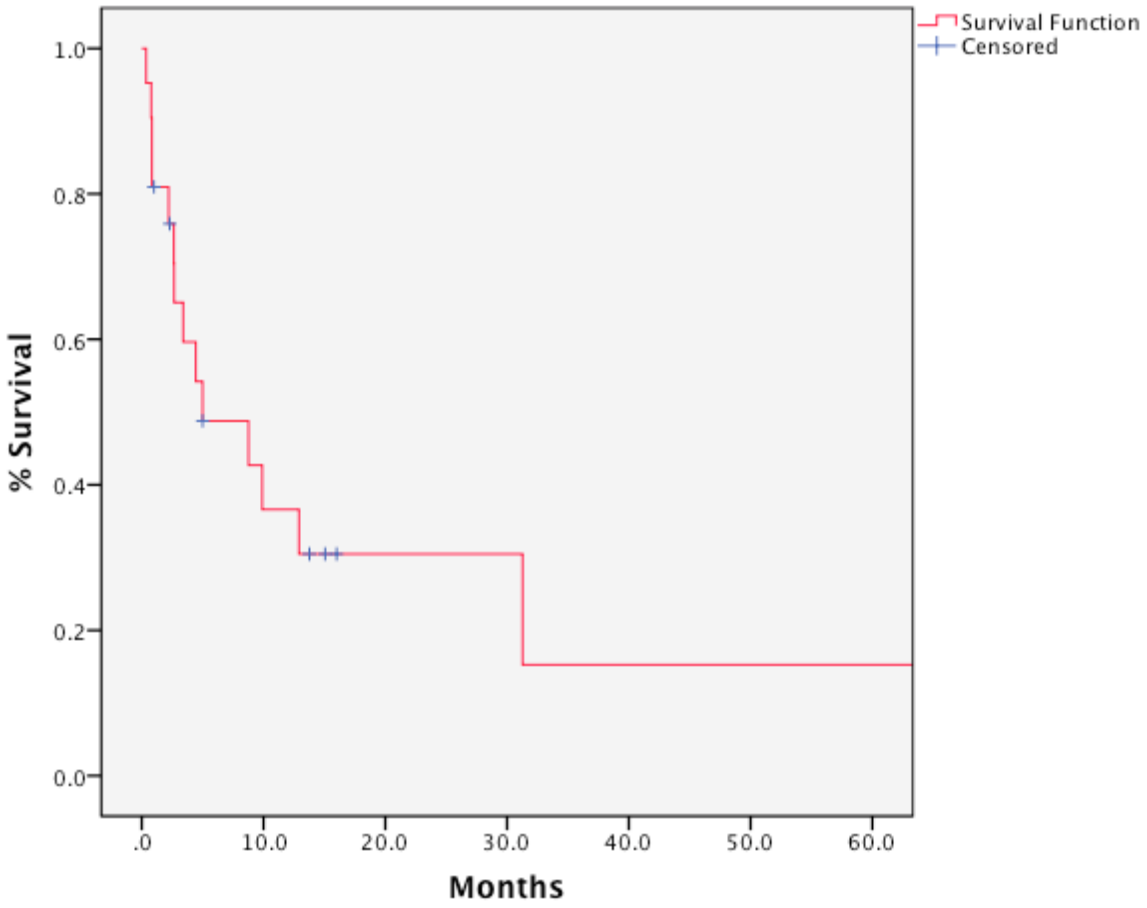
Carcinomatous and sarcomatoid elements were separately dissected from each tumor. Whole exome sequencing was separately performed on matched DNA samples comprising normal tissue, carcinomatous, and sarcomatoid components of each primary tumor. A summary of sequencing metrics is shown in Table 2. Normal, carcinomatous, and sarcomatoid components were sequenced to a mean depth of 135, 177, and 171 independent reads per targeted base in the exome. There was no correlation of median coverage in either carcinomatous or sarcomatoid regions with somatic mutation count ($R^2 < 0.01$, p=0.99 and $R^2<0.01$, p=0.94, respectively), suggesting sufficient depth of coverage for complete mutation calling. Median tumor purity was estimated at 62% (range 33%-82%) for the sarcomatoid and 46% (range 18%-75%) for the carcinomatous components (p=0.024 for difference in purity between components by Mann-Whitney U test). Variation in tumor purity between regions may be related to differences in the tumor microenvironment. Overall, the tumor purity in both components is similar to that observed in other cohorts such as TCGA (median 54%, range 18-87%).(22)

Table 1: Clinical Characteristics (n = 21)

Gender	Male	15 (71.4%)
	Female	6 (28.5%)
Age (years)	Mean	60.57 ± 12.3
	Median	61
Tumor size (cm)	Mean	11.27 ± 4.3
	Median	10
Tumor stage	T1	2 (9.5%)
	T2	2 (9.5%)
	T3	15 (71.4%)
	T4	2 (9.5%)
N stage	0	6 (76.2%)
	1+	15 (23.8%)
M stage	0	7 (33.3%)
	1	14 (66.6%)
% Sarcomatoid	Mean	46.9% ± 26.9%
	Median	50%

Means are shown with standard deviations

Figure 1:



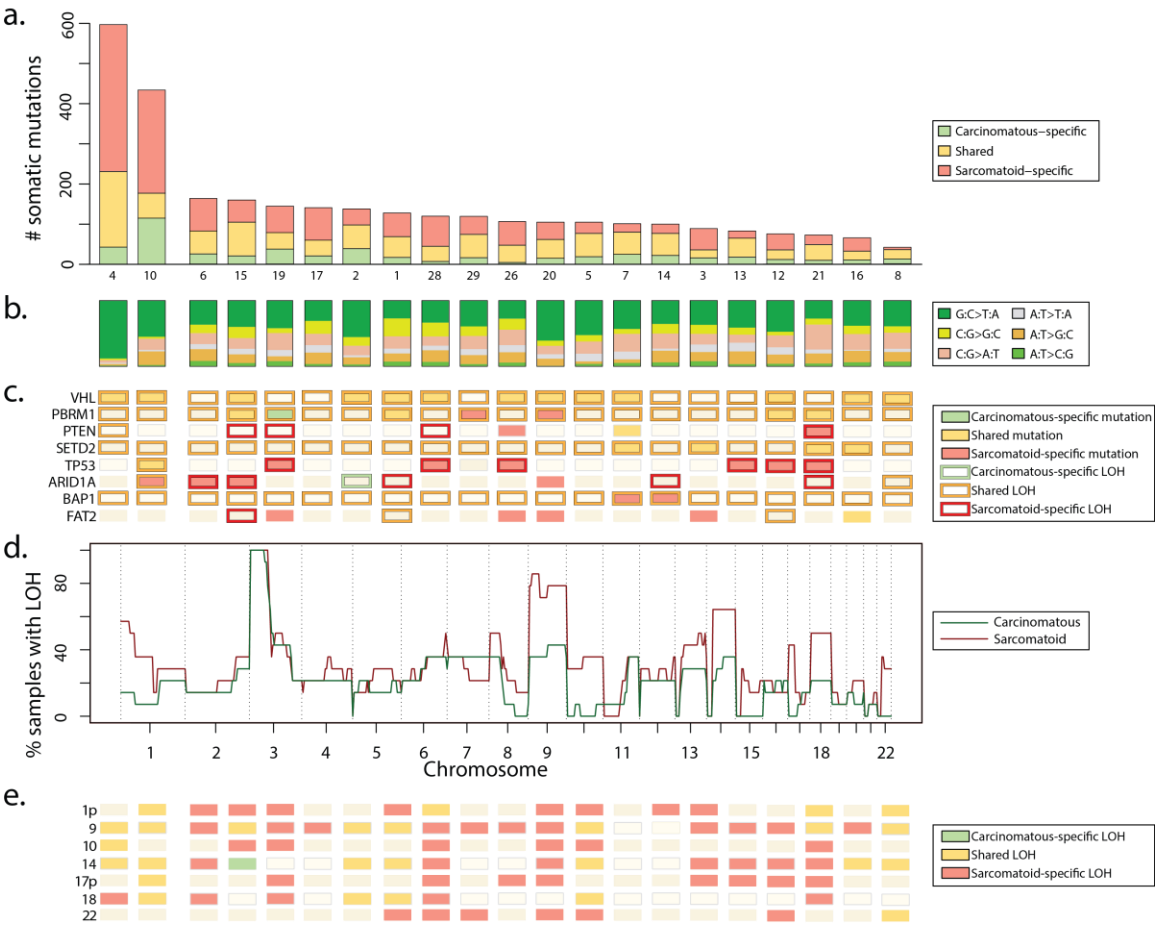
Patient Disease-Specific Survival

Table 2: Sequence Summary Statistics

Origin	Normal	Carcinomatous	Sarcomatoid
Number of samples	21	21	21
# of reads (M)	132	210	204
Median independent reads	109	150	142
Mean independent reads	135	177	171
% on genome	87.39%	87.29%	85.95%
% on target	70.47%	58.81%	57.96%
% of targeted bases $\geq$ 20 reads	94.08%	95.56%	94.76%
Mean error rate	0.71%	0.92%	0.94%

Landscape of mutation burden. Somatic single nucleotide variants (SSNVs) and chromosome segments showing loss of heterozygosity (LOH) were called in each tumor as described in the Methods. In 19 tumors, the mean total number of somatic single nucleotide variants (SSNVs), including both shared and component-specific, was 108 ± 33 (range 41 to 163; Figure 2a). The other two tumors were >5 standard deviation outliers in both tumor components, with a total mutation burden of 597 in one tumor and 434 SSNVs in the other. These two tumors also had a mutational signature characteristic of mismatch repair (MMR) deficiency, with an abundance of C:G>T:A transitions and a paucity of A:T>C:G, A:T>T:A, and C:G>G:C transversions(33) (Figure 2b). These tumors were considered to have hypermutation based on prior definitions (34, 35) and evidence of mismatch repair deficiency. Consistent with this, one hypermutated tumor had a heterozygous truncating mutation at R389 in *MSH2* in both carcinomatous and sarcomatoid elements with sarcomatoid-specific LOH at this locus as well as a sarcomatoid-specific heterozygous E1085K mutation in *POLE*. In contrast, there were no examples of hypermutation in the ccRCCs studied by TCGA (no tumor with more than 128 SSNVs)(22, 35) (p=0.002) (Figure 3). Neither of these samples had a germline alteration in MMR genes (*MSH2-6, MLH1, MLH3, PMS1-2, PSMP3, POLE*). In the TCGA cohorot, somatic alterations of MMR genes were infrequently observed(21/424 tumors, 5.0% overall), but there were no examples of hypermutation.(22) Both hypermutated tumors had homo/hemizygous VHL SSNVs, making misclassification unlikely, suggesting that these hypermutated tumors may be more likely to develop sarcomatoid features. As these

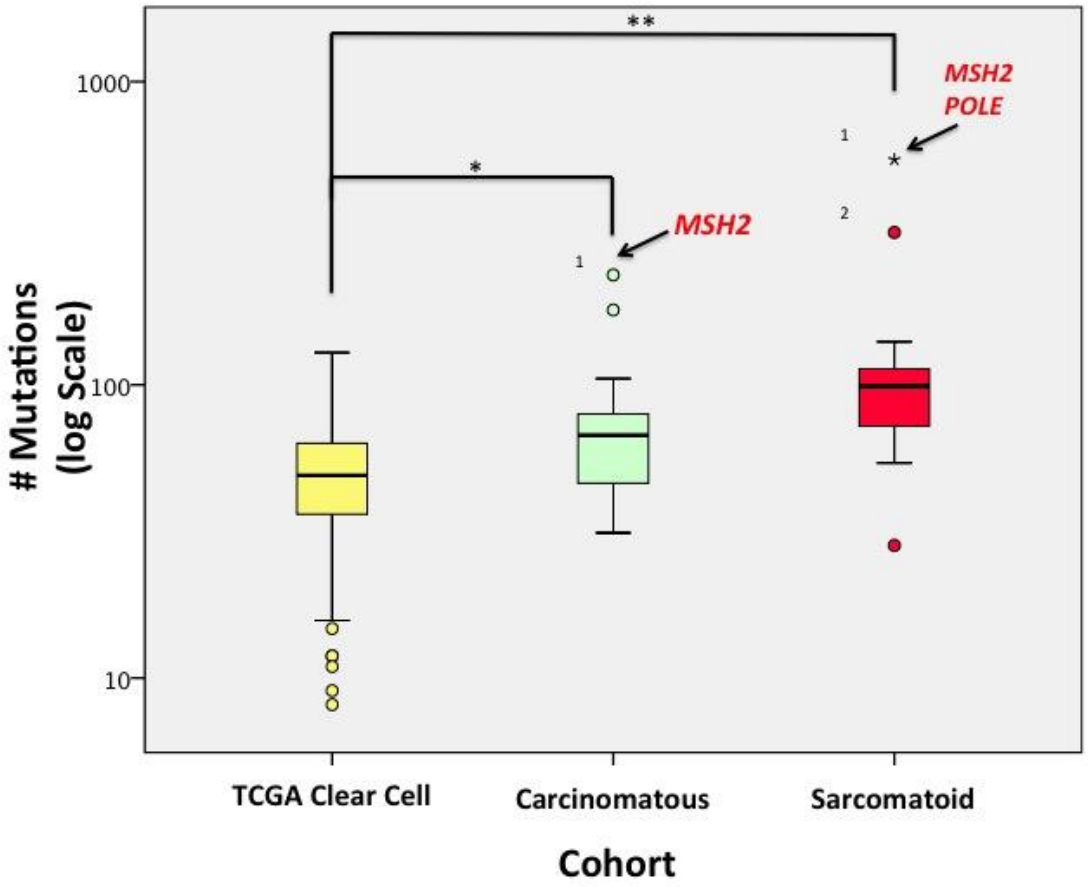
Figure 2:



Somatic mutations in 21 renal tumors with sarcomatoid features. a. Somatic mutation counts in 21 tumors by tumor component. Sample IDs labeled on bottom axis. b. Somatic mutation pattern by single nucleotide change. c. Presence of somatic mutations and LOH for significantly mutated and genes of interest. d. Frequency of LOH events by chromosome region in the carcinomatous (green) and sarcomatoid (red) tumor components for 14 non-hypermutated tumors with complete genome-wide LOH data. e. Presence of LOH in chromosomal segments with significant sarcomatoid-specific LOH.

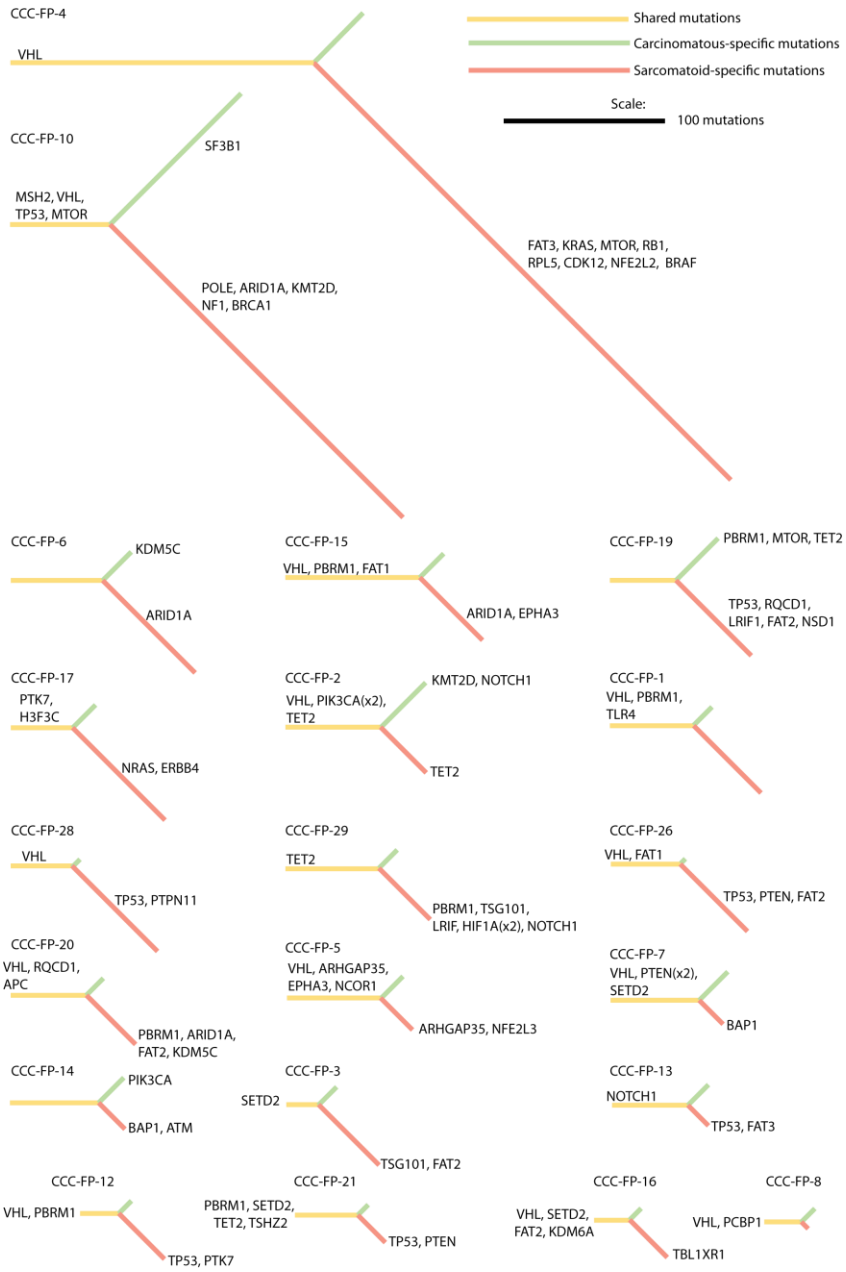
hypermutated tumors may differ from the others biologically, these were separately analyzed. Cancer genes with somatic mutations in these tumors are shown in Figure 4.

Figure 3:



Number of somatic mutations (Log Scale) in each cohort (TCGA Clear Cell/KIRC(22)) and clear cell and sarcomatoid component of our tumor cohort. Note hypermutated outliers (sample 1 and 2), one of which had a shared and a sarcomatoid-specific *MSH2* and *POLE* mutation, respectively. The median number of mutations was greater for both the carcinomatous and sarcomatoid components compared to the TCGA clear cell cohort (* and ** Mann-Whitney U test, p= 0.005 and <0.001, respectively)

Figure 4:



Phylogenetic trees of 21 ccRCC tumors. Branch and trunk lengths correspond to the number of somatic mutations in each tumor component, including shared, carcinomatous-specific, and sarcomatoid-specific mutations. Mutations in previously described ccRCC genes, Pan-Cancer genes, other recurrently mutated genes (*TSG101, RQCD1, LRIF1, PTK7*, and *FAT* family), and MMR genes in hypermutated samples are shown. Sample IDs labeled at top left of each phylogenetic tree.

Common origin of carcinomatous and sarcomatoid elements. Among the non-hypermutated tumors, sarcomatoid and carcinomatous elements shared a mean of 45/108 (41.7%) SSNVs, providing unequivocal evidence that these elements arise from a common cell of origin that bears many somatic mutations. The known cancer genes (using the Pan-Cancer gene set(36)) that most frequently shared somatic mutations in both carcinomatous and sarcomatous elements were *VHL* (SSNV+LOH in both elements in 11/19 tumors), *PBRM1* (SSNV+LOH in both elements in 4/19 tumors) and *SETD2* (SSNV+LOH in both elements in 4/19 tumors) (Figure 3c, Table 3). Moreover, these three genes are linked to one another on chromosome 3p; this segment shows LOH in every tumor. Somatic SSNVs in these three genes and LOH of 3p are hallmarks of ccRCC (22), and the evidence that these mutations predate the split of carcinomatous and sarcomatoid elements provides strong evidence that these tumors initially arise as ccRCC.

Table 3: Gene burden analysis of somatic mutations shared by carcinomatous and sarcomatoid elements of 19 non-hypermutated tumors.

Gene	Known ccRCC driver	Pan-Cancer gene	Coding size (bp)	# LOF or damaging missense mutations	P value	# other non-synonymous mutations	# silent mutations
VHL	Yes	Yes	645	11	**2.96E-32**	0	0
SETD2	Yes	Yes	7716	4	**2.60E-6**	0	0
PBRM1	Yes	Yes	4837	3	**1.98E-5**	1	0
PTEN	Yes	Yes	1221	2	**1.00E-4**	0	0
UBXN7	No	No	1481	2	1.51E-04	0	0
KLHL7	No	No	1772	2	1.65E-04	0	0
GBA	No	No	1622	2	1.75E-04	0	0
PIK3CA	Yes	Yes	3227	2	**5.39E-4**	0	0
USP53	No	No	3237	2	5.42E-04	0	0
NEDD4	No	No	3766	2	9.48E-04	0	0
TLR4	No	Yes	2523	1	**0.029**	0	0
TSHZ2	No	Yes	3107	1	**0.035**	0	0
KDM6A	Yes	Yes	4235	1	**0.042**	0	0

Significance threshold was 0.05 for known Pan-Cancer genes and 2.5×10^{-6} for all other genes. P-values in bold values indicate genes surpassing significance threshold. LOF, loss of function mutations- nonsense, splice site, or small insertion/deletions; damaging missense mutations- mutations at highly conserved positions or predicted as damaging by PolyPhen-2.

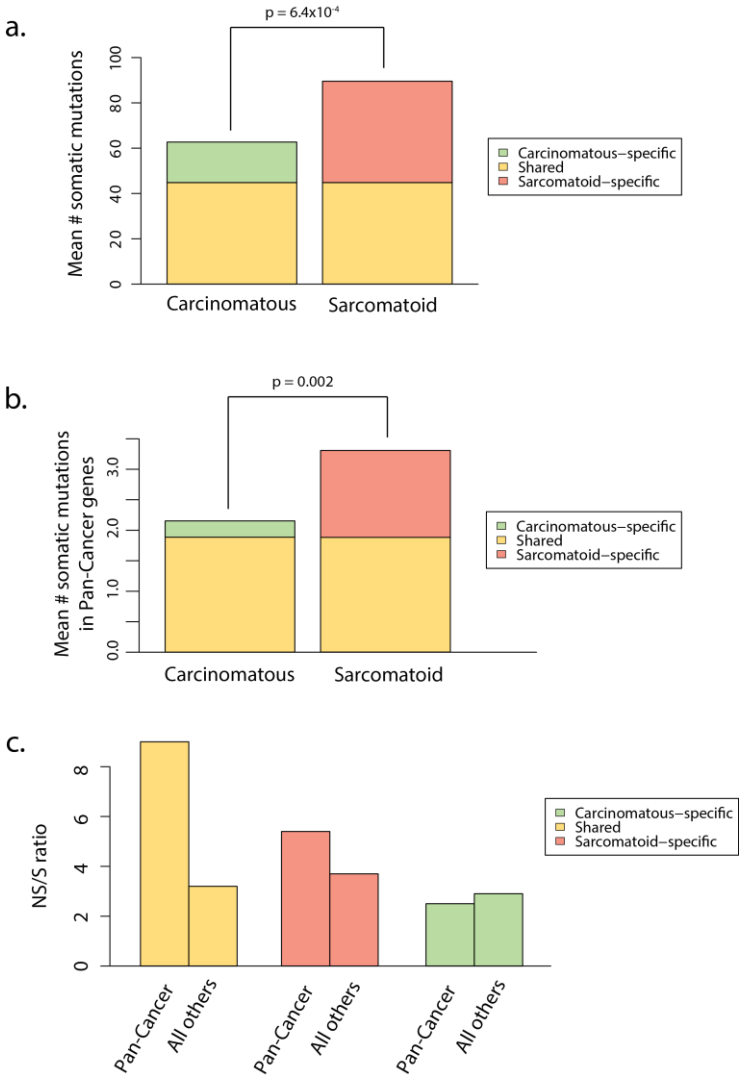
Increased burden of cancer driver mutations in sarcomatoid elements. Among somatic mutations that were specific to either sarcomatoid or carcinomatous elements, sarcomatoid components had a significantly higher burden of unique SSNVs (mean 45 vs. 18 SSNVs per tumor, $p=6.2x10^{-4}$, Figure 5a). Similarly, sarcomatoid components had nearly twice the length of element-specific somatic LOH (median 913 Mb vs 460 Mb, $p < 0.05$, Figure 2d).

The minor allele frequencies (MAFs) of component-specific SSNVs were significantly lower than those of shared SSNVs. Among carcinomatous components, the median MAFs were 15.3% vs. 21.4% for component-specific and shared SSNVs, respectively ($p<2.2x10^{-16}$ by Mann-Whitney U-test). Similarly, in sarcomatoid elements the median MAFs were 19.4% vs. 27.0% for component-specific and shared SSNVs, respectively ($p<2.2x10^{-16}$ by Mann-Whitney U-test). These findings are consistent with many component-specific mutations arising after clonal lineage separation.

Component-specific, non-synonymous SSNVs in known cancer genes were significantly more frequent in sarcomatoid than carcinomatous elements (respectively 27 total, mean 1.4 SSNVs per tumor, vs. 5 total, mean 0.26; p=0.002 by Wilcoxon signed rank test, Figure 5b). Non-synonymous somatic mutations in known cancer genes also occurred more often than expected by chance in sarcomatoid elements ($p = 1.7x10^{-6}$) but not carcinomatous elements ($p = 0.08$). Consistent with this, the ratio of non-synonymous/synonymous (NS/S) SSNVs was 5.4 for cancer genes vs. 3.7 for other genes in sarcomatoid elements, while the NS/S ratio was not elevated amongst Pan-cancer genes compared to other genes

in carcinomatous elements (2.5 vs 2.9) (Figure 5c). These findings lend further support to the evolution of sarcomatoid elements from carcinomatous elements by acquisition of new somatic mutations in cancer drivers rather than simply sampling issues from a heterogenous tumor.

Figure 5:



Comparison of somatic mutations in carcinomatous and sarcomatoid elements. a. Mean number of somatic mutations by tumor component for the 19 non-hypermutated tumors. 41.7% of all mutations were shared between tumor components. Sarcomatoid regions had a significantly higher number of component-specific mutations (mean 45 vs. 18, $p = 6.2x10^{-4}$ by Wilcoxon signed-rank text). b. Mean number of non-synonymous somatic mutations in known Pan-Cancer genes by tumor component. Sarcomatoid regions had a significantly higher number of component-specific mutations (1.42 vs 0.26, p = 0.002 by Wilcoxon signed-rank test). c. Ratio of non-synonymous to synonymous mutations in known Pan-Cancer genes by tumor component

Sarcomatoid-specific mutations in *TP53*, *ARID1A* and *BAP1*. Among sarcomatoid-specific SSNVs in known cancer genes, the frequency of mutation in *TP53* was remarkable (Table 4). There were no SSNVs or segments of LOH involving TP53 among carcinomatous regions in these tumors. In contrast, six sarcomatoid elements acquired bi-allelic *TP53* mutations (six NS SSNVs that were all homo/hemizygous via LOH), an event highly unlikely to occur by chance ($p=5.47x10^{-17}$, Figure 2c). Sarcomatoid-specific mutations also occurred in other cancer driver genes, including two mutations in *BAP1* and three in *ARID1A*. With the exception of one *ARID1A* mutation, all were accompanied by LOH ($p = 3.24x10^{-5}$ and $p=1.54x10^{-5}$ for presence of SSNV and LOH in *ARID1A* and *BAP1*, respectively). Interestingly, all biallelic *TP53*, *ARID1A,* and *BAP1* mutations were mutually exclusive (probability of 7.8% occurring by chance, direct calculation), suggesting that these SSNVs may represent alternative pathways towards sarcomatoid transformation (Figure 2c). Consistent with this interpretation, mutual exclusivity of *ARID1A* and *TP53* is commonly observed in ovarian and endometrial malignancies.(37, 38) Similarly in the clear cell TCGA cohort, mutations in *TP53* (9 tumors) and *ARID1A* (14 tumors) were mutually exclusive. Both genes have been shown to form a complex that regulates transcription of *CDKN1A* and *SMAD4*.(37) As either mutation alone may be sufficient alone to promote tumorigenesis via a common pathway, it may be unnecessary for the cell to have concurrent mutations in both genes.

Table 4: Gene burden analysis of sarcomatoid-specific mutations in 19 non-hypermutated tumors.

Gene	Known ccRCC driver	Pan-Cancer gene	Coding size (bp)	# LOF or damaging missense mutations	P value	# other non-synonymous mutations	# silent mutations
TP53	Yes	Yes	1192	6	**2.28E-14**	0	0
FAT2	No	No	13047	4	4.84E-05	0	0
ARID1A	Yes	Yes	6878	3	**5.04E-05**	0	0
TSG101	No	No	1183	2	6.94E-05	0	0
PTEN	Yes	Yes	1221	2	**9.80E-05**	0	0
BAP1	Yes	Yes	2207	2	**2.41E-04**	0	1
HIF1A	No	No	2496	2	**3.05E-04**	0	0
LRIF1	No	No	2314	2	5.23E-04	0	0
PBRM1	Yes	Yes	4837	2	**2.44E-03**	0	0
NRAS	No	Yes	1797	1	**8.55E-03**	0	0
PTPN11	No	Yes	1559	1	**0.020**	0	0
TBL1XR1	No	Yes	2969	1	**0.023**	0	0
EPHA3	No	Yes	3955	1	**0.043**	0	0
ERBB4	No	Yes	574	1	**0.044**	0	0

Significance threshold was 0.05 for known Pan-Cancer genes and 2.5×10^{-6} for all other genes. P-values in bold values indicate genes surpassing significance threshold. LOF, loss of function mutations- nonsense, splice site, or small insertion/deletions; damaging missense mutations- mutations at highly conserved positions or predicted as damaging by PolyPhen-2.

Table 5: Gene burden analysis of the carcinomatous-specific mutations in the 19 non-hypermutated tumors.

Gene	Known ccRCC driver	Pan-Cancer gene	Coding size (bp)	# LOF or damaging missense mutations	P value	# other non-synonymous mutations	# silent mutations
RLIM	No	No	1878	2	5.20E-05	0	0
PRPF8	No	No	7050	2	7.08E-04	0	0
KDM5C	Yes	Yes	4709	1	**0.025**	0	0
MTOR	Yes	Yes	7707	1	**0.035**	0	0
PBRM1	Yes	Yes	4837	1	**0.041**	0	0

Significance threshold was 0.05 for known Pan-Cancer genes and 2.5×10^{-6} for all other genes. P-values in bold values indicate genes surpassing significance threshold. LOF, loss of function mutations- nonsense, splice site, or small insertion/deletions; damaging missense mutations- mutations at highly conserved positions or predicted as damaging by PolyPhen-2.

Among tumors with genome-wide LOH data in both components, several chromosomes showed recurrent segments of sarcomatoid-specific LOH that were unlikely to have occurred by chance. These included segments on chromosome 1p (57%, all including *ARID1A*, q=0.030); chr. 9 (86%, all including *CDKN2A*, q=0.007), chr. 10 (36%, all including *PTEN*, q=0.108) chr. 14 (64%, q=0.108), chr. 17p (43%, all including *TP53*, q=0.030), chr. 18 (50%, q=0.188) and chr. 22 (29% tumors, q=0.210) (Figure 2d, 2e).

We also sought other genes with SSNVs that occurred more often than expected by chance on either lineage (Table 4, 5). FAT atypical cadherin 2 (*FAT2*) was the second most frequently mutated gene in the sarcomatoid-specific gene burden analysis (4 SSNVs) and the top mutated gene not previously implicated in ccRCC ($p = 4.84x10^{-5}$). Six other genes not previously implicated in ccRCC harbored sarcomatoid-specific SSNVs in two tumors each, some in segments of LOH. These genes include two additional *FAT* genes, *FAT1* and *FAT3*, as well as tumor susceptibility 101 (*TSG101*), ligand dependent nuclear receptor interacting factor 1 (*LRIF1*), required for cell differentiation 1 homolog (*RQCD1*), and protein tyrosine kinase 7 (*PTK7*). Details of these and other SSNVs in driver genes are shown in Table 6. Phylogenetic trees of all tumors, by tumor component and including mutated cancer genes, are shown in Figure 4.

Table 6: Non-synonymous somatic mutations in select genes

Sample	Samples with mutation*	Gene	Chr	Pos	# ref sarc	# non-ref sarc	# ref carc	# non-ref carc	AA change	# non-conserved species†	PolyPhen-2 score‡	LOH*
CCC-FP-10§	CS	*TP53*	17	7578406	39	5	70	10	R175H	2	0.999(D)	CS
CCC-FP-12	S	*TP53*	17	7579355	61	32	130	0	L111P	5	1(D)	S
CCC-FP-13	S	*TP53*	17	7577081	121	32	248	0	E286G	1	0.995(D)	S
CCC-FP-19	S	*TP53*	17	7578413	8	7	61	0	V173M	3	1(D)	S
CCC-FP-21	S	*TP53*	17	7578211	138	19	271	1	R213Q	1	1(D)	S
CCC-FP-26	S	*TP53*	17	7577551	55	8	112	0	M246in_frame_del	NA	NA	S
CCC-FP-28	S	*TP53*	17	7578538	94	99	185	0	N131I	5	1(D)	S
CCC-FP-1	CS	*VHL*	3	10188253	118	42	128	46	Q132	NA	NA	CS
CCC-FP-10§	CS	*VHL*	3	10191512	123	15	96	47	L169	NA	NA	CS
CCC-FP-12	CS	*VHL*	3	10183725	13	8	18	3	S65X	NA	NA	CS
CCC-FP-15	CS	*VHL*	3	10188251	139	180	191	62	Q132fs	NA	NA	CS
CCC-FP-16	CS	*VHL*	3	10183703	23	9	14	1	R58fs	NA	NA	CS
CCC-FP-2	CS	*VHL*	3	10191476	29	62	98	35	T157fs	NA	NA	CS
CCC-FP-20	CS	*VHL*	3	10183800	9	6	20	5	N90in_frame_del	NA	NA	CS
CCC-FP-26	CS	*VHL*	3	10188245	99	90	194	14	V130L	0	0.979(D)	CS

CCC-FP-28	CS	*VHL*	3	10183762	9	8	15	10	C77X	NA	NA	CS
CCC-FP-4§	CS	*VHL*	3	10183861	6	6	1	5	H110fs	NA	NA	CS
CCC-FP-5	CS	*VHL*	3	10188295	152	87	167	81	P146fs	NA	NA	CS
CCC-FP-7	CS	*VHL*	3	10183797	5	3	11	7	L89H	4	1(D)	CS
CCC-FP-8	CS	*VHL*	3	10188210	331	39	214	80	L118R	1	1(D)	CS
CCC-FP-10§	S	*ARID1A*	1	27101611	20	9	34	3	Q1414fs	NA	NA	CS
CCC-FP-15	S	*ARID1A*	1	27099030	80	88	228	0	S1149X	NA	NA	S
CCC-FP-20	S	*ARID1A*	1	27105553	99	10	170	3	R1505X	NA	NA	-
CCC-FP-6	S	*ARID1A*	1	27099947	39	88	128	12	R1276X	NA	NA	S
CCC-FP-14	S	*BAP1*	3	52436820	80	13	133	1	E653fs	NA	NA	CS
CCC-FP-7	S	*BAP1*	3	52443860	63	14	42	0	P12L	1	0.997(D)	CS
CCC-FP-1	CS	*PBRM1*	3	52682418	194	122	171	79	I252T	2	0.901(P)	CS
CCC-FP-12	CS	*PBRM1*	3	52613113	76	56	125	14	V1139fs	NA	NA	CS
CCC-FP-15	CS	*PBRM1*	3	52643672	49	45	48	13	E742X, P741P	NA	NA	CS
CCC-FP-19	C	*PBRM1*	3	52582176	146	0	230	29	A1551fs	NA	NA	CS
CCC-FP-20	S	*PBRM1*	3	52668672	276	54	230	2	K416fs	NA	NA	CS
CCC-FP-21	CS	*PBRM1*	3	52713619	55	32	121	31	R37fs	NA	NA	CS
CCC-FP-29	S	*PBRM1*	3	52643477	214	27	239	0	I807fs	NA	NA	CS

CCC-FP-21	S	*PTEN*	10	89711913	145	62	417	0	Y177X	NA	NA	CS
CCC-FP-26	S	*PTEN*	10	89624292	161	117	384	1	D22fs	NA	NA	CS
CCC-FP-7	CS	*PTEN*	10	89720677	120	43	180	68	N276K	0	1(D)	CS
CCC-FP-7	CS	*PTEN*	10	89717724	239	33	338	48	C250in_frame_ins	NA	NA	CS
CCC-FP-16	CS	*SETD2*	3	47164856	51	23	60	16	R424X	NA	NA	CS
CCC-FP-21	CS	*SETD2*	3	47147612	27	17	48	16	2 bp upstream of exon 6	NA	NA	CS
CCC-FP-3	CS	*SETD2*	3	47129632	299	36	311	63	Q1750X	NA	NA	CS
CCC-FP-7	CS	*SETD2*	3	47162421	21	11	39	29	L1235fs	NA	NA	CS
CCC-FP-15	CS	*FAT1*	4	187628245	17	11	30	8	V913M	35	0.143(B)	-
CCC-FP-26	CS	*FAT1*	4	187539105	193	99	364	45	K2879E	13	0.005(B)	S
CCC-FP-16	CS	*FAT2*	5	150945470	12	2	15	11	L1008fs	NA	NA	-
CCC-FP-19	S	*FAT2*	5	150914030	18	9	51	0	V3122in_frame_del	NA	NA	-
CCC-FP-20	S	*FAT2*	5	150947852	22	5	24	1	R214Q	10	0.958(D)	-
CCC-FP-26	S	*FAT2*	5	150911450	30	6	24	0	R3170H	6	0.996(D)	-
CCC-FP-3	S	*FAT2*	5	150923703	168	32	158	0	G2329R	0	1(D)	-
CCC-FP-4[§]	S	*FAT3*	11	92085964	73	9	100	0	L229P	2	1(D)	S
CCC-FP-13	S	*FAT3*	11	92620268	218	11	203	0	G4347A	7	0.792(D)	-
CCC-FP-12	S	*PTK7*	6	43098336	99	40	154	1	A250D	1	1(D)	-

CCC-FP-17	CS	*PTK7*	6	43111336	133	18	25	4	E613D	0	0.999(D)	S
CCC-FP-19	S	*RQCD1*	2	219457100	36	14	80	0	R205H	0	1(D)	-
CCC-FP-20	CS	*RQCD1*	2	219457091	311	5	158	1	T202M	0	1(D)	S
CCC-FP-19	S	*LRIF1*	1	111494368	20	12	66	1	D380fs	NA	NA	-
CCC-FP-29	S	*LRIF1*	1	111492586	272	52	285	0	R586G	1	0.992(D)	-
CCC-FP-3	S	*TSG101*	11	18503287	385	10	343	0	L324F, Y325N	0	0.999(D)	-
CCC-FP-29	S	*TSG101*	11	18531165	211	52	299	0	F135fs	NA	NA	-

*S = mutation/LOH is exclusive to sarcomatoid region; C = exclusive to carcinomatous region; CS = shared by sarcomatoid and carcinomatous regions.

†Number of vertebrate species with substitutions at this position.

‡PolyPhen-2 score (classification: D- probably damaging; P- possibly damaging; B- benign).

§Hypermutated tumor

Discussion

The presence of sarcomatoid features has long been recognized as an extremely poor prognostic factor in kidney cancer. However, until now, a genetic basis of sarcomatoid transformation has remained largely unknown. Our finding that 43% of somatic mutations are shared between carcinomatous and sarcomatous elements provides conclusive evidence that these elements arise from a common clonal ancestor. Despite the suggestion of a shared clonal origin, there has previously been little evidence that the sarcomatoid component arose in a process of de-differentiation from a pre-existing carcinomatous component.(16) The data herein provides very strong evidence of a carcinomatous origin. The most frequently mutated cancer drivers that are shared by carcinomatous and sarcomatous elements are the genes that are characteristically mutated in ccRCC. Second, the burden of new SSNVs in known cancer drivers is more than 5-fold higher on the sarcomatoid than carcinomatous regions. Third, there is a highly significant burden of sarcomatoid-specific mutation of *TP53*, implicating a specific gene in development of sarcomatoid elements, along with recurrent mutations and/or segments of LOH affecting other known cancer genes. These findings support a pathogenic sequence in which somatic mutations occurring in a ccRCC drives de-differention to a sarcomatoid state. Importantly, the finding of highly significant sarcomatoid-specific mutation signals is inconsistent with the observed differences being the result of simple heterogeneity within tumors, in which case differences between carcinomatous and sarcomatoid elements would be expected to be stochastic.

The high frequency of homozygous *TP53* mutations in sarcomatoid elements was striking. These were found in 6 of 19 (31.5%) non-hypermutated tumors. *TP53* mutations are otherwise rare in ccRCC ((22, 39). For example among 395 ccRCCs reported by TCGA that do not have sarcomatoid elements, only 6 of 395 (1.5%) have *TP53* mutation ($P = 3 \times 10^{-6}$, Fisher exact test, odds ratio 29), and only two of these are in segments of LOH (Figure 6). These findings are consistent with prior work from Oda and colleagues who performed a candidate gene study of 14 tumors with sarcomatoid transformation and noted a higher incidence of *TP53* mutations in the sarcomatoid region using immunohistochemistry and Sanger sequencing.(40). *TP53* alterations may link the EMT pathway to sarcomatoid transformation, as p53 loss can reduce expression of miR-200c, which contributes to EMT.(41) Additionally, one hypermutated tumor contained an R175H alteration in *TP*53, a known gain-of-function mutation that results in upregulation of TWIST1, an important EMT transcriptional regulator (42). In another form of renal cancer, Wilms tumors, loss of *TP53* similarly leads to histologic changes (anaplasia) and a poor prognosis.(43, 44)

We identified somatic mutations in genes that are characteristic for ccRCC, including *VHL, PBRM1, SETD2, PTEN, ARID1A*, and *BAP1.* Notably, all *ARID1A* and *BAP1* SSNVs were exclusive to sarcomatoid regions, all but one were in segments of LOH, and were mutually exclusive with each other and with *TP53* mutations. Deficiency of *ARID1A* and *BAP1* has been associated with worse prognosis, higher tumor grade, and a higher incidence of sarcomatoid histology.(45, 46) However, many other tumors show only LOH at these loci,

suggesting these events may be permissive of but insufficient for sarcomatoid transformation.

The incidence of *VHL* SSNV's was 57.9% (11/19 tumors). Additionally, all tumors had LOH of chromosome 3p. Consistent with its role as an early event in tumorigenesis(47, 48), all *VHL* mutations were shared in carcinomatous and sarcomatoid elements. *VHL* alterations (mutation and hypermethylation) have been considered the hallmark of ccRCC.(49, 50) For centrally reviewed ccRCC, the incidence of *VHL* mutation is as high as 81.3%(51), which suggests that our cohort had a lower incidence of *VHL* mutation.(39, 51) Similarly, it has been shown that wild-type *VHL* ccRCC has a more aggressive phenotype, (52, 53) perhaps related to an increased propensity for sarcomatoid transformation.

Mutations in genes not implicated in ccRCC may be relevant to sarcomatoid transformation. Sarcomatoid-specific mutations in *FAT2* were found in 5 tumors (including one in a hypermutated tumor). Mutations in other members of the FAT family, including *FAT1* and *FAT3*, were in 2 tumors each. In total, 7 of 19 non-hypermutated tumors (36.8%) in our cohort contained mutations FAT family genes. In contrast, FAT family mutations were rarely found in ccRCC in TCGA (5.8%, Figure 6). FAT proteins play multiple roles in cell adhesion, motility, polarity, signaling, and proliferation, and mutations are implicated in a variety of cancers(54-56). Loss of *FAT1* has been shown to promote WNT signaling, a critical mediator of EMT(57). Further exploration of these genes in sarcomatoid elements will be required to assess the significance of the role of FAT family genes.

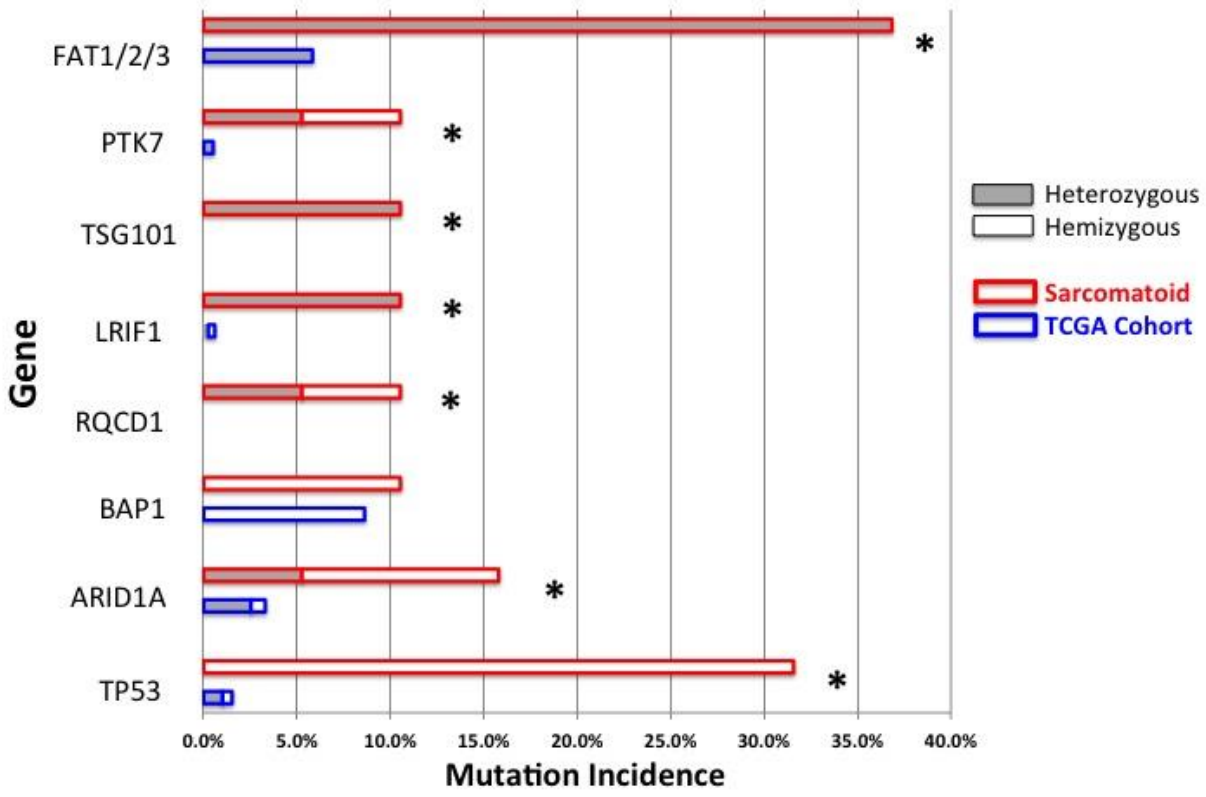
Mutations in several other genes were of interest, but will also require larger numbers of samples to assess significance. All mutations in these genes are listed in Table 6. Two tumors had sarcomatoid-specific mutations in *TSG101*, a member of the ESCRT-I complex involved in ubiquitinated protein trafficking(58) and a known breast cancer tumor suppressor (58, 59). These include a frameshift mutation and a dinucleotide mutation in a highly conserved position in the "steadiness box" domain, responsible for cellular auto-regulation of *TSG101*.(58) Both overexpression and deficiency of *TSG101* have been linked to tumorigenesis and cell cycle abnormalities.(58) Two mutations, one sarcomatoid-specific and one shared with sarcomatoid-specific LOH, both in highly conserved amino acid positions, were present in *PTK7*, a tyrosine kinase regulator of cell motility, adhesion, polarity, and WNT signaling(60). Both over and under-expression of *PTK7* is implicated in a variety of cancers(61). In human pluripotent stem cells it appears to be a marker of EMT(62). Lastly, two mutations each were present in *RQCD1* and *LRIF1*, both retinoic-acid receptor transcriptional cofactors(63, 64). RQCD1 is a component of the DRF complex containing RAR and ATF-2,(63) as well as a component of the CCR4-NOT complex involved in miRNA-mediated gene silencing(65). While *RQCD1* mutations were extremely rare in TCGA (2/424), they were exclusively found in sarcomatoid tumors ($p= 0.0020$). *RQCD1* has previously been implicated in breast cancer and melanoma(65, 66). LRIF1 is a nuclear matrix-associated protein that may repress ligand-dependent transcriptional activation by RARα (64). Several of these genes were enriched in specimens listed as having sarcomatoid transformation in the TCGA dataset (Figure 6).

Effective systemic therapy for individuals with sarcomatoid renal tumors is an unmet need in oncology. *TP53* and *ARID1A* are among potential sarcomatoid-specific targets for which therapeutics are in development (67, 68). As drugs that mitigate effects of mutation in these genes enter clinical trials, their study in sarcomatoid renal tumors may be warranted. Similarly, hypermutability in sarcomatoid tumors also has implications for treatment of both tumor components. Loss of key mismatch repair genes can sensitize tumors to radiation and some types of chemotherapy.(69) These tumors have also shown sensitivity to immunotherapy for several cancer types, perhaps due to the increased burden of novel epitopes. (70, 71). Lastly, PD-1 and PDL-1 expression has also recently been found to be greater in tumors with sarcomatoid features,(72) raising the possibility that these tumors may be responsive to immune checkpoint inhibitor immunotherapy.

This study is the first comprehensive genomic assessment of the process of sarcomatoid transformation in ccRCC using next generation sequencing techniques. Other strengths of our study include the limitation of our cohort to a single subtype of RCC (clear cell carcinoma) rather than an admixture of differing subtypes with different underlying biology and the collaboration of three independent pathologists for confirmation of sarcomatoid transformation. The latter point is particularly vital as there is no objective marker of sarcomatoid transformation. Our study was limited by the use of only a single site within each tumor component per patient, which limited our ability to perform a more comprehensive phylogenetic assessment. Our comparisons to the TCGA cohort

may be confounded by differences in bioinformatics pipelines and somatic mutation filtering criteria. Additionally, TCGA samples did not undergo central pathology review for sarcomatoid histology. The presence or absence of sarcomatoid histology in these tumors should therefore be in question: tumors marked as having sarcomatoid features may not have truly contained sarcomatoid histology or may have contained mixed components.

Figure 6:



	TP53	*ARID1A*	*BAP1*	*RQCD1*	*LRIF1*	*TSG101*	*PTK7*	*FAT1/2/3*
TCGA KIRC Overall	6/395	13/395	34/395	0/395	2/395	0/395	2/395	23/395
Sarcomatoid Component	6/19	3/19	2/19	2/19	2/19	2/19	2/19	7/19
p value	***3.10E-06***	***3.15E-02***	6.76E-01	***2.00E-03***	***1.14E-02***	***2.00E-03***	***1.14E-02***	***1.40E-04***

Comparison of frequency of somatic mutation of most frequently mutated in sarcomatoid regions (red outline) (n=19) and that observed in non-sarcomatoid clear cell tumors in the TCGA (blue outline) (n=395). Heterozygous alterations are shaded gray while hemizygous alterations are in white. Fisher's Exact test was used for comparisons. * denotes a p value of <0.05. *TP53, ARID1A, RQCD1, LRIF1, TSG101, PTK7,* and *FAT1/2/3* were all significantly increased in the sarcomatoid components.

References:

1. Shuch B, *et al.* (2009) Cytoreductive nephrectomy for kidney cancer with sarcomatoid histology--is up-front resection indicated and, if not, is it avoidable? *J Urol* 182(5):2164-2171.
2. de Peralta-Venturina M, *et al.* (2001) Sarcomatoid differentiation in renal cell carcinoma: a study of 101 cases. *Am J Surg Pathol* 25(3):275-284.
3. Shuch B, *et al.* (2015) Understanding pathologic variants of renal cell carcinoma: distilling therapeutic opportunities from biologic complexity. *Eur Urol* 67(1):85-97.
4. Shuch B, *et al.* (2012) Impact of pathological tumour characteristics in patients with sarcomatoid renal cell carcinoma. *BJU international* 109(11):1600-1606.
5. Zhang BY, *et al.* (2015) A novel prognostic model for patients with sarcomatoid renal cell carcinoma. *BJU Int* 115(3):405-411.
6. Delahunt B, *et al.* (2013) The International Society of Urological Pathology (ISUP) grading system for renal cell carcinoma and other prognostic parameters. *Am J Surg Pathol* 37(10):1490-1504.
7. Cheville JC, *et al.* (2004) Sarcomatoid renal cell carcinoma: an examination of underlying histologic subtype and an analysis of associations with patient outcome. *Am J Surg Pathol* 28(4):435-441.
8. Merrill MM, *et al.* (2015) Clinically nonmetastatic renal cell carcinoma with sarcomatoid dedifferentiation: Natural history and outcomes after surgical resection with curative intent. *Urol Oncol*.
9. Shuch B, Bratslavsky G, Linehan WM, & Srinivasan R (2012) Sarcomatoid renal cell carcinoma: a comprehensive review of the biology and current treatment strategies. *Oncologist* 17(1):46-54.
10. Golshayan AR, *et al.* (2009) Metastatic sarcomatoid renal cell carcinoma treated with vascular endothelial growth factor-targeted therapy. *J Clin Oncol* 27(2):235-241.
11. Voss MH, *et al.* (2014) Treatment outcome with mTOR inhibitors for metastatic renal cell carcinoma with nonclear and sarcomatoid histologies. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 25(3):663-668.
12. Nanus DM, Garino A, Milowsky MI, Larkin M, & Dutcher JP (2004) Active chemotherapy for sarcomatoid and rapidly progressing renal cell carcinoma. *Cancer* 101(7):1545-1551.
13. Farrow GM, Harrison EG, Jr., & Utz DC (1968) Sarcomas and sarcomatoid and mixed malignant tumors of the kidney in adults. *Cancer* 22(1,2, and 3):545-563.
14. Anani W, Amin M, Pantanowitz L, & Parwani AV (2014) A series of collision tumors in the genitourinary tract with a review of the literature. *Pathology, research and practice* 210(4):217-223.
15. Darai-Ramqvist E, Nilsonne G, Flores-Staino C, Hjerpe A, & Dobra K (2013) Microenvironment-Dependent Phenotypic Changes in a SCID Mouse Model for Malignant Mesothelioma. *Frontiers in oncology* 3:203.

16. Delahunt B (1999) Sarcomatoid renal carcinoma: the final common dedifferentiation pathway of renal epithelial malignancies. *Pathology* 31(3):185-190.
17. Jones TD*, et al.* (2005) Clonal Divergence and Genetic Heterogeneity in Clear Cell Renal Cell Carcinomas with Sarcomatoid Transformation. *Cancer* 104(6):1195-1203.
18. Oda H & Machinami R (1993) Sarcomatoid renal cell carcinoma. A study of its proliferative activity. *Cancer* 71(7):2292-2298.
19. Kuroiwa K, Konomoto T, Kumazawa J, Naito S, & Tsuneyoshi M (2001) Cell proliferative activity and expression of cell-cell adhesion factors (E-cadherin, alpha-, beta-, and gamma-catenin, and p120) in sarcomatoid renal cell carcinoma. *J Surg Oncol* 77(2):123-131.
20. Shuch B*, et al.* (2010) Histologic evaluation of metastases in renal cell carcinoma with sarcomatoid transformation and its implications for systemic therapy. *Cancer* 116(3):616-624.
21. Conant JL, Peng Z, Evans MF, Naud S, & Cooper K (2011) Sarcomatoid renal cell carcinoma is an example of epithelial--mesenchymal transition. *J Clin Pathol* 64(12):1088-1092.
22. Cancer Genome Atlas Research N (2013) Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499(7456):43-49.
23. Guo G*, et al.* (2012) Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. *Nature genetics* 44(1):17-19.
24. Beroukhim R*, et al.* (2009) Patterns of gene expression and copy-number alterations in von-hippel lindau disease-associated and sporadic clear cell carcinoma of the kidney. *Cancer research* 69(11):4674-4681.
25. Goh G*, et al.* (2014) Recurrent activating mutation in PRKACA in cortisol-producing adrenal tumors. *Nat Genet* 46(6):613-617.
26. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.
27. Cibulskis K*, et al.* (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology* 31(3):213-219.
28. Zhao S*, et al.* (2013) Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* 110(8):2916-2921.
29. Genomes Project C*, et al.* (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56-65.
30. Adzhubei I, Jordan DM, & Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]* Chapter 7:Unit7 20.
31. Pleasance ED*, et al.* (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463(7278):191-196.
32. Cerami E*, et al.* (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* 2(5):401-404.

33. Alexandrov LB, *et al.* (2013) Signatures of mutational processes in human cancer. *Nature* 500(7463):415-421.
34. Pritchard CC, *et al.* (2014) Complex MSH2 and MSH6 mutations in hypermutated microsatellite unstable advanced prostate cancer. *Nature communications* 5:4988.
35. Davis CF, *et al.* (2014) The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* 26(3):319-330.
36. Kandoth C, *et al.* (2013) Mutational landscape and significance across 12 major cancer types. *Nature* 502(7471):333-339.
37. Guan B, Wang TL, & Shih Ie M (2011) ARID1A, a factor that promotes formation of SWI/SNF-mediated chromatin remodeling, is a tumor suppressor in gynecologic cancers. *Cancer Res* 71(21):6718-6727.
38. Bosse T, *et al.* (2013) Loss of ARID1A expression and its relationship with PI3K-Akt pathway alterations, TP53 and microsatellite instability in endometrial cancer. *Mod Pathol* 26(11):1525-1535.
39. Sato Y, *et al.* (2013) Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet* 45(8):860-867.
40. Oda H, Nakatsuru Y, & Ishikawa T (1995) Mutations of the p53 gene and p53 protein overexpression are associated with sarcomatoid transformation in renal cell carcinomas. *Cancer Res* 55(3):658-662.
41. Chang CJ, *et al.* (2011) p53 regulates epithelial-mesenchymal transition and stem cell properties through modulating miRNAs. *Nat Cell Biol* 13(3):317-323.
42. Kogan-Sakin I, *et al.* (2011) Mutant p53(R175H) upregulates Twist1 expression and promotes epithelial-mesenchymal transition in immortalized prostate cells. *Cell death and differentiation* 18(2):271-281.
43. Bardeesy N, Beckwith JB, & Pelletier J (1995) Clonal expansion and attenuated apoptosis in Wilms' tumors are associated with p53 gene mutations. *Cancer Res* 55(2):215-219.
44. Maschietto M, *et al.* (2014) TP53 mutational status is a potential marker for risk stratification in Wilms tumour with diffuse anaplasia. *PLoS One* 9(10):e109924.
45. Pena-Llopis S, *et al.* (2012) BAP1 loss defines a new class of renal cell carcinoma. *Nat Genet* 44(7):751-759.
46. Kapur P, *et al.* (2013) Effects on survival of BAP1 and PBRM1 mutations in sporadic clear-cell renal-cell carcinoma: a retrospective analysis with independent validation. *Lancet Oncol* 14(2):159-167.
47. Gerlinger M, *et al.* (2014) Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature genetics* 46(3):225-233.
48. Gerlinger M, *et al.* (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England journal of medicine* 366(10):883-892.
49. Gnarra JR, *et al.* (1994) Mutations of the VHL tumour suppressor gene in renal carcinoma. *Nat Genet* 7(1):85-90.

50. Herman JG, *et al.* (1994) Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* 91(21):9700-9704.
51. Nickerson ML, *et al.* (2008) Improved identification of von Hippel-Lindau gene alterations in clear cell renal tumors. *Clinical cancer research : an official journal of the American Association for Cancer Research* 14(15):4726-4734.
52. Patard JJ, *et al.* (2009) Absence of VHL gene alteration and high VEGF expression are associated with tumour aggressiveness and poor survival of renal-cell carcinoma. *Br J Cancer* 101(8):1417-1424.
53. Gordan JD, *et al.* (2008) HIF-alpha effects on c-Myc distinguish two subtypes of sporadic VHL-deficient clear cell renal carcinoma. *Cancer cell* 14(6):435-446.
54. Furukawa T, *et al.* (2015) Whole exome sequencing reveals recurrent mutations in BRCA2 and FAT genes in acinar cell carcinomas of the pancreas. *Scientific reports* 5:8829.
55. Morris LG, Ramaswami D, & Chan TA (2013) The FAT epidemic: a gene family frequently mutated across multiple human cancer types. *Cell cycle* 12(7):1011-1012.
56. Katoh M (2012) Function and cancer genomics of FAT family genes (review). *International journal of oncology* 41(6):1913-1918.
57. Morris LG, *et al.* (2013) Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. *Nature genetics* 45(3):253-261.
58. Feng GH, Lih CJ, & Cohen SN (2000) TSG101 protein steady-state level is regulated posttranslationally by an evolutionarily conserved COOH-terminal sequence. *Cancer research* 60(6):1736-1741.
59. Sundquist WI, *et al.* (2004) Ubiquitin recognition by the human TSG101 protein. *Molecular cell* 13(6):783-789.
60. Peradziryi H, Tolwinski NS, & Borchers A (2012) The many roles of PTK7: a versatile regulator of cell-cell communication. *Archives of biochemistry and biophysics* 524(1):71-76.
61. Golubkov VS, *et al.* (2014) Protein-tyrosine pseudokinase 7 (PTK7) directs cancer cell motility and metastasis. *The Journal of biological chemistry* 289(35):24238-24249.
62. Chan DN, Azghadi SF, Feng J, & Lowry WE (2012) PTK7 marks the first human developmental EMT in vitro. *PloS one* 7(11):e50432.
63. Hiroi N, *et al.* (2002) Mammalian Rcd1 is a novel transcriptional cofactor that mediates retinoic acid-induced cell differentiation. *The EMBO journal* 21(19):5235-5244.
64. Li HJ, Haque ZK, Chen A, & Mendelsohn M (2007) RIF-1, a novel nuclear receptor corepressor that associates with the nuclear matrix. *Journal of cellular biochemistry* 102(4):1021-1035.
65. Wong SQ, *et al.* (2015) Whole exome sequencing identifies a recurrent RQCD1 P131L mutation in cutaneous melanoma. *Oncotarget* 6(2):1115-1127.

66. Ajiro M, *et al.* (2009) Involvement of RQCD1 overexpression, a novel cancer-testis antigen, in the Akt pathway in breast cancer cells. *International journal of oncology* 35(4):673-681.
67. Hong B, van den Heuvel AP, Prabhu VV, Zhang S, & El-Deiry WS (2014) Targeting tumor suppressor p53 for cancer therapy: strategies, challenges and opportunities. *Current drug targets* 15(1):80-89.
68. Bitler BG, *et al.* (2015) Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers. *Nat Med* 21(3):231-238.
69. Martin SA, Lord CJ, & Ashworth A (2010) Therapeutic targeting of the DNA mismatch repair pathway. *Clin Cancer Res* 16(21):5107-5113.
70. Rizvi NA, *et al.* (2015) Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348(6230):124-128.
71. Le DT, *et al.* (2015) PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* 372(26):2509-2520.
72. Joseph RW, *et al.* (2015) PD-1 and PD-L1 Expression in Renal Cell Carcinoma with Sarcomatoid Differentiation. *Cancer Immunol Res*.