RESPONSE LETTER

-- Ref1.1 – Significance about the results of MET --

Reviewer Comment	The authors have focused on MET and produced some data that did not provide further advances to what we have known so far on the role of MET in type I pRCC.
Author Response	We thank the reviewer for expressing concerns about our results on MET. Indeed, MET has been known to be the central driver in type I pRCC for decades. However, most of the analysis focuses on the coding region only. Moreover, the majority of type I pRCC patients in the TCGA study do not carry any missense mutation in MET. ~20% patients show significantly higher MET expression yet are completely silent in MET alterations, without evidence for missense mutations, alternative splicing and copy number amplification. Using a more integrated approach, particularly focusing on non-coding, we are able to provide hints for alternative mechanisms to MET dysfunction in type I pRCC. In particular: 1. Our study is the first one that comprehensively looks into the non-coding regions of pRCC. Doing so allow us to tackle an open question in the field of cancer genomics, whether whole genome sequencing adds additional value over exome sequencing. Recent studies in whole genome sequencing suggest active roles of non-coding mutations in cancer. Well-known examples include TERT promoter mutations in urothelial carcinoma and enhancer hijacking in CNS tumors (REF). However, the debate of WGS versus WES remains unsettled and significances of many non-coding alterations stay unknown. In this study, we investigate the functional roles of non-coding alterations in pRCC. We find excessive non-coding mutations at the promoter and regulatory regions of MET. Given the critical role MET plays in pRCC and some MET-driven samples are completely silent in terms of alterations of MET, we believe this mutation hotspot is possibly linked with pRCC molecular etiologies. Accordingly, we have revised the manuscript to better explain the significance of our findings. 2. During our revision, we find the activation of a cryptic promoter in the second intron of MET causes the alternative mRNA isoform described in the original TCGA study. This event has been observed in several other cancers included CML and some GI (gastrointestinal tract)

	cancers. We provide an explanation for the alternative MET transcription isoform in pRCC. Further more, we linked the usage of this cryptic promoter with the methylation change that is often seen in pRCC. We added this new analysis in the revised manuscript. 3. We find more somatic mutations in an extended WXS set, further completing the MET mutation spectrum of the TCGA study.
Excerpt From Revised Manuscript	Our study is the first one that comprehensively looks into the noncoding regions of pRCC. Doing so allow us to tackle an open question in the field of cancer genomics, whether whole genome sequencing adds additional value over whole exome sequencing. The TCGA study has identified a MET alternate transcript as a driver event (3). However, the etiology of this new isoform is unknown. We identified this alternate transcript results from the usage of a cryptic promoter from an L1 element (Figure 1A), likely due to a local loss of methylation (REF). This event was reported in several other cancer types (REF). To test its relationship with methylation, we found a closet probe (cg06985664, ~3kb downstream) on the methylation array shows marginally statistically significant (p=0.055, one-sided rank-sum test). Additionally, this event is associated with methylation group 1 (odds ration (OR)= 4.54, p<0.041), indicating genome-wide methylation dysfunction. This association is stronger in type 2 pRCC and it shows a significant association with the C2b cluster (OR= 17.5, p<0.007).

-- Ref1.2 – Non-coding analysis power--

Reviewer Comment	The non-coding analysis did not show significant findings, likely due to the small cohort size and the heterogeneous nature (cohort (n=32) included 19 type I pRCC, 6 type II pRCC, and 7 unclassified).
Author Response	We agree with the referee that our traditional statistical test power is limited. However, in terms of the size of data, our 35 WGS samples (with three newly added samples, see below) have more reads than >1,000 WES samples. Unlike the traditional statistics test by looking at a single (or a few) feature in the cohorts, we have the ability to obtain a large number of features (in forms of genomic regions) about each sample. This forms a high-dimensional scenario (p>>n, "short, fat data") commonly seen in big data analysis nowadays. As the referee points out, low cohort number limits our statistical power. But we instead conduct a comprehensive and unbiased examination of the entire genome for each sample. The great amount of data we acquire from every single sample greatly boosts our analyses. The impactful non-coding alterations we identify are in fact filtered out from thousands of changes in the entire genome and we have high confidence that they have truly high impacts. In our study, we are able to show a significant amount of samples carry impactful mutations in noncoding regions and conduct some coarse recurrence tests. Our analysis is the first exploration of pRCC non-coding regions and provides meaningful insights of

Shantao 2/24/2017 10:54 AM
Deleted: affected

	pRCC. This hopefully will spark some research ideas and interests in noncoding regions of pRCC.
	1. The non-coding mutation hot spots indeed carry excessive and impactful mutations. We segment the genome based on functional annotation (FunSeq). Then we try to find highly recurrent mutations in annotated regions. These three mutation hotspots have extremely high mutation rate in our cohort. The hotspots span from 7 to 50kb, each with 6-to-7 mutations observed in 35 samples (~150,000 non-coding mutations in total). Therefore, the local mutation rate is roughly 5-to-20 times higher than average. We explain our approaches better in the revised manuscript. 2. We leverage the existing knowledge in coding regions and complete the picture of cancer genomes with our non-coding analyses. All three hotspots are tightly linked with coding genes that are biologically associated with pRCC. Mutations in these regions could have high impacts. Unfortunately, non-coding regions are largely overlooked in the previous studies of pRCC. Our study is the first one that looks into these regions that make up to 98% of the genome. Although we were not able to perform fine-scale tests for these mutation hotspots due to sample size, we hope our analyses will spark interests and encourage researcher to further explore the possible biological impacts of these events. 3. In our revision process, we reviewed the WGS samples and added three more WGS samples into our cohort, reaching a final size of 35. We also want to point out that, because WGS covers more than 50 times more regions than WES, additional three samples add more reads than 100 WES samples. Thus this added data greatly boosted our data-driven analyses, for example, signature and mutation landscape study.
Excerpt From Revised Manuscript	Expanding our scope from coding to non-coding and use FunSeq to group SNVs by functional elements, we found several potentially significant non-coding mutation hotspots relevant to tumorigenesis throughout the entire genome.

-- Ref1.3 – Implications of NEAT1 mutations--

Reviewer Comment	This reviewer was very intrigued by the NEAT1 finding, which deserves more work to elucidate its importance and could be the highlight of this paper. Can we use NETA1 promoter mutation to classify pRCC and what are the associated transcriptomic signature?
Author Response	Recurrent mutations in NEAT1 are indeed intriguing. NEAT1 is a non-coding RNA thus will be missed by whole exome sequence. It was overlooked in previous studies of pRCC. Our study is the first one on NEAT1 in pRCC. We show a mutation hotspot in NEAT1 and mutations are linked with higher expression of

NEAT1, presumably due to the dysfunction of gene regulation region, and worse survival of patients. As the referee suggested, we did additional work on NEAT1 in the revised manuscript.

Although lacking WGS data to find genomic alteration, we found NEAT1 is overexpressed in 5% of the TCGA ccRCC cohort. NEAT1 higher expression is significantly associated with shorter overall survival time (median OS: 36 months versus 77 months). NEAT1 is tightly co-expressed with MALAT1 in both pRCC and ccRCC. COSMIC annotates MALAT1 as consensus cancer driver in pediatric RCC and lung cancer.

The referee raised an interesting point about expression signature. NEAT1 mutations seem to be associated with RNAseq cluster 3 but do not reach statistical significance (p>0.05), likely due to our small WGS sample size. We used two other cluster methods for mRNA expression (Firehose, Broad Institute). Although, again, our NEAT1 mutation status is not significantly associated with mRNA clusters, we find NEAT1 as a marker gene. Its expression level significantly (FDR<0.05) differs in different clusters. Besides, NEAT1 tightly co-expressed with MALAT1, which is a known cancer gene. COSMIC (REF) annotates it with pediatric RCCs.

We expect with a larger cohort, the statistical significance we get will be strengthened. As an active participant of the currently ongoing PCAWG (PanCancer Analysis of Whole Genomes), we quickly looked into NEAT1 mutations in the high quality PCAWG RCC dataset. 21/144(14.58%) of the samples carry mutations in NEAT1, a frequency agrees with the one from our cohort.

WE add a new section and a supplemental figure to reflect the discussions above and our new analysis results of NEAT1.

Excerpt From Revised Manuscript	To further investigate the role of NEAT1 in RCCs, we found NEAT1 overexpression is significantly associated with shorted overall survival in TCGA ccRCC cohort (p=0.0132, Fig SXX). Moreover, MALAT1, another noticeable lncRNA in cancer, is tightly co-expressed with NEAT1 in both pRCC and ccRCC (Spearman's correlation: 0.79 and 0.87 respectively). Catalogue of Somatic Mutations in Cancer (COSMIC) (REF) annotates MALAT1 as cancer consensus gene, associating it with pediatric RCCs and lung cancer. Overexpression of MALAT1 is reported to be associated with cancer progression (REF).

-- Ref1.4 –Significance of mutation spectra & landscape--

Reviewer Comment	The findings on mutation spectra and defects in chromatin remodeling affecting mutation landscape are of moderate interest.
Author Response	We appreciate the referee for raising concerns about the significance of the mutation spectra and landscape analysis.

Several recent landmark pan-cancer studies lead to the wide recognition of significance and great research interests in cancer mutational processes (REF). DNA mutation is one of the important driving forces of cancer development. Understanding the underlying processes and affecting factors that generate the mutations is vital in cancer studies.

As the referee points out earlier, pRCC is very heterogeneous, especially the type II. TCGA study shows several subgroups of pRCC while we still observe great variation in subgroups. A key aim of our study is to better understand this heterogeneity. Studying the heterogeneity in the fundamental processes that generate the mutations brings about great research excitement. Mutation spectra elucidate diversified mutation processes in pRCC. In our study, we identify several factors (methylation, APOBEC, chromatin remodeling defects etc.) that play important roles in tumorigenesis. This helps better characterize and understand pRCC in terms of variations in mutagenesis, tumor evolution, and molecular etiologies. It also has potential clinical implications. For instance, mutation burden has important predictive value on immune therapy response. In the era of great advances of immune therapy, we feel research on mutation landscape in pRCC has the potential to facilitate clinical decisions.

Moreover, WGS provides unique advantages to mutation spectra and landscape analyses. First, WGS produces more than an order more SNVs in the sample, which greatly powers the analyses. Second, most of the noncoding regions are not subject to strong selection and thus minimizes the selection confounding effects. Last, WGS does not need the exon capture preprocess step before sequencing. Therefore, WGS avoids related bias and technical artifacts.

In the revised manuscript we add discussions to better explain the significance of this part of the study.

| Excerpt From Revised Manuscript | WGS provides many times more SNVs compared to WXS, and noncoding SNSs are less constrained by selection pressure. Thus it gives us a great opportunity to look into the high-level landscape of mutations in pRCC…..

…We identified mutation rate dispersion of C-to-T in CpG motif contributes the most to the inter-sample mutation spectra variations….In our study, we observed C-to-Ts in CpG are enriched in coding regions, which indicates they have higher functional impacts in cancer genome.

Researchers found tumors with DNA mismatch repair deficiency response better to PD-1 |

-- Ref1.5 – Individual evolution trees --

Reviewer Comment	The WGS analysis is somewhat descriptive. With the wealth of this dataset, the author shall attempt to generate individual pRCC evolution trees of these 32 cases.
Author Response	We thank the referee for the suggestion. WGS indeed provides a large amount of SNVs (> 16,000 in total, which equals more than 100 WES samples), making our dataset suitable for evolution analysis. In the revision, we build individual evolutionary trees for 35 cases. This is the first study inferring tumor evolutionary trees using large number of SNVs from WGS in pRCC. Benefited from a large number of SNVs, the tree construction becomes more accurate and reveals more details.

We identify four major types of trees, based on topology and the number of subclones identified: linear (no branch) with fewer subclones, linear with more subclones, shortly branched and branched with distal subclones. We find some interesting correlations of tree topology and cancer subtypes. Trees with short branches are enriched in type 1 RCC (p<0.002) while trees with long branches or linear tree with more subclones occurs more often in type 2. The evolution tree structure might reflect some intriguing cancer biological processes in RCC. Type I RCCs are enriched in more homogenous tree structures. Subclones exist but they are not divergent. Whereas type II RCCs have more subclones and show greater degree of divergence. This is in line with the knowledge that type II RCCs are more heterogeneous.

A major feature of RCCs is heterogeneity. Our evolutionary tree study demonstrates the tumors also have distinct evolutionary profiles. Evolutionary tree reflects the level of tumor heterogeneity, tumor development process and accumulating of mutations. Our results help understand the origin and mechanism of heterogeneity.

We add a new figure panel in figure 4 and corresponding results/discussion sections for the new evolution tree analysis. Also, we show 35 individual trees, one for every single sample, in the supplement figures. |
| Excerpt From Revised Manuscript | With the richness of SNVs in WGS samples, we further tackle the mutational process heterogeneity of pRCC by constructing individual evolutionary trees for 35 tumors (Figure SXX). Three trees have largest population faction <0.5 (likely due to low mutation number, high sequence error and/or high heterogeneity) and thus are |

Shantao 2/24/2017 8:26 PM
Deleted: .

excluded from downstream analysis. We could further classify the trees into four types based on topology (Figure 4A, 4B): (1) no branch, fewer subclones (10, 32.3%), (2) short branches (12, 37.5%), (3) no branch, more subclones (5, 15.6%) and (4) long branches (5, 15.6%). Both (3) and (4) show significant clonal evolution, indicated by long mutations distances between populations.

Short branch type is significantly enriched in Type I pRCC (p<0.011, two-tailed fisher exact test) while the more heterogeneous types: long branches and no branch, more subclones type are significantly depleted in Type I (p < 0.0034, two-tailed fisher exact test). This indicates type I tumors are more homogenous and show less complex evolution features compared to type II and unclassified samples.

-- Ref1.6 – Minor --

Reviewer Comment	a) line 173, please add reference b) line 258, based on available clinical trials, there is almost certain that c-MET inhibitor has no role in type II pRCC, which needs to be rephrased. c) line 278-283, will expand pending further analysis
Author Response	We thank the referee for pointing out these issues. In the revision, we a) added reference to support higher mutation rate of C-to-T in methylated CpGs. (T.R. Waters, P.F. Swann Thymine-DNA glycosylase and G to A transition mutations at CpG sites Mutat. Res., 462 (2000), pp. 137–147) b) We understand the evidence for MET inhibitor efficacy in type II pRCC is not very solid. But we do not think we can completely rule out the possibility of patients with rs11762213 response to it. Especially we should notice that only 3-5% of the RCC patients carry rs1176113 and currently existing trials did not assess it as a biomarker. Meanwhile, 46% type II patients were found to have MET alterations (Albiges et al, 2014). Also a dual MET/VEGF inhibitor, Foretinib, was found to have high response rate in patients with germline alterations in MET (Choueiri et al., 2011). Last, although MET is not playing a central role in ccRCC, ccRCC that had failed with VEGFR TKIs response to Cabozantinib well (Choueiri et al,, 2015; METEOR) Designation of papillary subtype is based on light microscopy. There is often discordance between histologic subtype and molecular classification. The S1500 SWOG trial (https://clinicaltrials.gov/ct2/show/NCT02761057) includes all patients with papillary RCC and the specific integrated biomarkers investigate pathologic subtype and MET alterations with response rate. This is the purpose of the clinical trial, which has just started. The question is important to the NCI GU Steering committee and is why it is funded by the NCI.

	Rs11762213 genotyping is reliable and the cost is low, we suggested this might become a biomarker to predict the patient response for MET inhibitor. Nonetheless, to avoid confusion, we rephrased the sentence to emphasize on using rs11762213 as a potential biomarker pending more studies. c) We expanded the section of NEAT1. See REF1.3.
Excerpt From Revised Manuscript	b) Also, rs11762213 might become a biomarker for predicting patient response to MET inhibitors.

-- Ref2.1 – Molecular mechanisms of rs11762213 --

Reviewer Comment	For the germline SNP rs11762213, it does not change protein sequence. If it really plays some role in cancer, it probably has regulatory function(s). However, the authors didn't observe changes in expression or protein abundance of MET. I am wondering what about the expression and protein abundance of MET in ccRCC where this SNP also is associated with prognosis. And what about genes that are next to MET in both pRCC and ccRCC if MET is unchanged?
Author Response	The referee raised an excellent question. The mechanism of rs11762213, a synonymous exonic SNP, remains still unsettled. A recent publication about rs11762213 by AA Hakimi et al. studied this in great details in ccRCC. They did not find any statistically significant change in MET expression patterns associated with this SNP. Also this SNP is not in strong linkage disequilibrium with other SNPs of interest in RCCs. Given this, we follow the referee's suggestion and look for expression changes in adjacent genes. However, there is no gene other than MET within 100kb both up- and downstream of the gene. Together with the central role of MET in pRCC, we think this SNP is likely to affect prognosis through MET, although the mechanism remains unknown. Since this is a germline SNP, it may affect tumor development, even at the very early stage. Such effects might be complicated and become cryptic during the tumor development and thus fail to be detected. Also this SNP might have affect the MET expression in nearby tissues and stimulate the tumor growth. AA Hakimi et al., were not able to get statistical significance on higher MET expression in normal tissue associated with rs11762213. We explored our dataset and unfortunately only found one sample

	with risk allele that also has normal tissue MET expression level. This impedes statistical testing. Finally, we also looked into the Genotype-Tissue Expression (GTEx) project, and did not find evidence for rs11762213 affecting gene expression. However, this could be due to low statistical power because of small normal kidney sample size in these studies and rs11762213 being a relatively rare SNP. In the revised manuscript, we better elaborate the current research status of rs11762213 and incorporate the discussions above.
Excerpt From Revised Manuscript	A previous study proposes it disrupt a putative enhancer and thus affect *MET* expression. However, researchers cannot find significant difference in *MET* expression in either tumor or normal tissues. We noticed there is other gene within 100 kb of this SNP. Given the significant role of *MET* in pRCC, we also think rs11762213 is affecting survival through *MET*, although the mechanism unknown.

-- Ref2.2 – DHS validity --

Reviewer Comment	The authors shall use caution when counting mutations in DHS sites when there is mutation in chromatin remodelers. The authors claimed mutations in chromatin remodelers can change the chromatin environment. If so, comparing number of mutations in DHS sites predicted from one cell line will particularly be problematic in patients with mutations in remodelers.
Author Response	The referee made an excellent observation. We certainly agree that, DHS regions called from a normal kidney cell line represent the open chromatin regions under normal, physiological condition. With chromatin remodeling dysfunction, the DHS regions are likely to shift in pRCC tumors. In fact we believe this is a very plausible explanation for mutation landscape changes since chromatin environment greatly affects DNA repair and replication. We admit the language we use in the manuscript causes confusion. In the revised manuscript, we use "open chromatin regions in normal kidney tissue" to accurately describe the nature of these DHS sites. Further more, to address the concern of referee, in the revision we find a better match for the normal kidney tissue under physiological condition. We substitute the cell line with eleven fetal kidney cortex tissues. Using kidney tissues better reflects the physiological, normal kidney tissue. Also gathering data from eleven tissues allows us minimizing experimental bias and variance. Last, since the eleven tissues come from different fetus with different genetic background, we are able to reduce individual variance and call conservative/stable DHS sites by taking the intersection of the DHS regions. Last, DHS regions are enriched with functional regions of

Shantao 2/18/2017 9:03 PM
Deleted: cells

Shantao 2/24/2017 11:00 AM
Deleted: I

Shantao 2/24/2017 11:01 AM
Deleted: ,

Shantao 2/24/2017 11:03 AM
Deleted: substitute

	genome, for example, essential genes. Therefore, a higher mutation burden in DHS regions might be deleterious for tumor. Nonsynonymous mutations in protein coding regions may also be antigenic. Recent studies have shown patients with higher and impactful mutation burden responsed better to immunotherapy. Thus this shift of mutation landscape may have clinical implications.
Excerpt From Revised Manuscript	To test this hypothesis, we tallied the number of mutations inside DNase I hypersensitive sites (DHS) in eleven normal fetal kidney cortex samples (The NIH Roadmap Epigenomics Mapping Consortium, REF), which represent the normal, physiological condition. …driven by an even stronger mutation rate increase in putative open chromatin regions in normal kidney tissues. This is likely because chromatin remodeling defects affect normal open chromatin environment and impede DNA repairing in these regions.

-- Ref2.3 – Figure 2A --

Reviewer Comment	Figure 2A is confusing. There are 3 proposed promoters and 4 SNVs in promoter, inconsistent with text. It's better to put this panel into Figure 1 rather than in Figure 2.
Author Response	We thank the reviewer for pointing the flaws in our figure preparation. We have fixed the promoter regions and put it into Figure 1.
Excerpt From Revised Manuscript	

-- Ref2.4 – Color key in Figure 4 --

Reviewer Comment	Color key should be added in Figure 4
Author Response	We thank the reviewer for pointing the flaws in our figure preparation. We have added color key in Figure 4
Excerpt From Revised Manuscript	

-- Ref3.1 – The significance of rs11762213 in pRCC--

Reviewer Comment	They looked at an exonic SNP in the MET gene among pure papillary RCC (rather than mixed RCC histologies done previously) and found marginally worse prognosis in type 2 pap RCC with the SNP. They argue that this may have clinical implications and that patients with the SNP may benefit from MET inhibitors. However, the association is not strong enough for it to matter clinically. A cost benefit analysis would be needed as well as an explanation of how it would impact management. The claim that it would select patients for MET inhibition is unsubtantiated. The authors link this SNP to a racial predisposition to developing papillary RCC ... but this is mostly speculation.

Author Response	We totally agree with the reviewer that there is a long path to translate scientific discoveries in the lab into clinical care. In this scientific research article, we have no intention to offer any suggestion for clinical practice changes. Cost-benefit analysis and many more studies are certainly needed before any change in patient management. We are afraid that they are beyond the scope of the article and *Plos Genetics*. 1. The two previous studies about rs11762213 were done on a mixed RCC cohort and a cohort entirely made up of TCGA ccRCC respectively. The mixed cohort was mostly ccRCC (78% in discovery cohort and 75% in validation cohort) due to the disease nature. The pRCC subset is apparently too small to run any subgroup analysis. Both of the studies were not able to prove rs11762213 predict prognosis in pRCC. Without proper subtype stratification, a plausible alternative is rs11762213 only predicts prognosis in ccRCC. In this manuscript, for the first time, we find that rs11762213 has predictive value in type 2 pRCC outcome. 2. The p-value indicates the chances that the null hypothesis is true. It is certainly impacted by the magnitude of the effects of the SNP. But, many other factors also greatly affect the p-value, for example, statistical power/sensitivity. In our case, the p-value is largely bounded by the small sample size. A "marginal" p-value does not necessarily mean the effect of the SNP on prognosis is small. 3. We were forming hypotheses and speculating about the etiologies and implications of rs11762213 in the discussion section. We agree with the reviewer that we should rewrite this part to better explain the implications of our study. Thus we revised the SNPs discussion in the manuscript.
Excerpt From Revised Manuscript	Not significantly mutated in ccRCC and type 2 pRCC, MET nonetheless seems to play a role in cancer development. This finding is potentially meaningful in clinical management of patients with the more aggressive type 2 pRCC. rs11762213 genotyping could become a reliable, low-cost risk stratification tool for these patients. Also, rs11762213 might become a biomarker for predicting patient response to MET inhibitors….. …This implies a possible effect of rs11762213 on pRCC incidence among African Americans that is worth further investigation.

-- Ref3.2 –Statistical significance--

Reviewer Comment	Their analysis of non-coding mutation hotspots was largely negative or statistically underpowered. They found mutations in the promoter region of NEAT1, a non-coding RNA, which were marginally associated with worse outcome. This is interesting but of minor significance.
Author Response	We understand the concern of the reviewer. However, we feel the recurrent mutations in NEAT1 are actually of great interest.

	First, NEAT1 is a non-coding RNA thus will be missed by whole exome sequence. It was overlooked in previous studies of pRCC. We conducted the first study of NEAT1 in pRCC. We show a mutation hotspot in NEAT1 and mutations are linked with higher expression of NEAT1, presumably due to the dysfunction of gene regulation region, and worse survival of patients. As the referee suggested, we did additional work on NEAT1 in the revised manuscript. Although lacking WGS data to find genomic alteration, we found NEAT1 is overexpressed in 5% of the TCGA ccRCC cohort. NEAT1 higher expression is significantly associated with shorter overall survival time (median OS: 36 months versus 77 months). NEAT1 is tightly co-expressed with MALAT1 in both pRCC and ccRCC, which is another noticeable lncRNA in cancer. The referee raised an interesting point about expression signature. NEAT1 mutations seem to be associated with RNAseq cluster 3 but do not reach statistical significance (p>0.05), likely due to our small WGS sample size. We used two other cluster methods for mRNA expression (Firehose, Broad Institute) and again, our NEAT1 status is not significantly associated with mRNA clusters. However, we find NEAT1 as a marker gene, its expression level significantly differs in different clusters. Besides, NEAT1 tightly co-expressed with MALAT1, which is a known cancer gene. We expect with a larger cohort, the statistical significance we get will be strengthened. As an active participant of the currently ongoing PCAWG (PanCancer Analysis of Whole Genomes), we quickly look into NEAT1 mutations in the high quality PCAWG RCC dataset. 21/144(14.58%) of the samples carry mutations in NEAT1, a frequency agrees with the one from our cohort. WE add a new section and a supplemental figure to reflect the discussions above and our new analysis results of NEAT1.
Excerpt From Revised Manuscript	NEAT1 is overexpressed in 5% ccRCC samples from the TCGA cohort. NEAT1 overexpression is significantly associated with shorted overall survival (Fig SXX). MALAT1, another noticeable lncRNA in cancer, is tightly co-expressed with NEAT1 in both pRCC and ccRCC (Spearman's correlation: 0.79 and 0.87 respectively). Catalogue of Somatic Mutations in Cancer (COSMIC) (REF) annotates MALAT1 as cancer consensus gene, associating it with pediatric RCCs and lung cancer. Overexpression of MALAT1 is reported to be associated with cancer progression (REF).

-- Ref3.3 – Interpretation of APOBEC--

Reviewer Comment	They found an APOBEC mutation signature in only 1 out of 155 cases. Given that APOBEC signatures are described in urothelial carcinoma, the authors then theorized that papillary RCC may be genomically similar to urothelial carcinoma ... and may potentially be managed similarly with chemotherapy and radiation therapy. This is a great leap of faith and logic (or illogic). Again, attesting to the paucity of actual positive findings.
Author Response	We thank the reviewer for expressing the concerns about our interpretation of APOBEC and the language we use here. pRCC is very heterogeneous, especially the type II. TCGA study shows several subgroups of pRCC and still we see large variation within subgroups. A key aim of our study is to better understand this heterogeneity. APOBEC mutagenesis shows both location (prefer single-strand DNA, for example around double strand break sites) and context (unique trinucleotide signature) preference. Therefore, in APOBEC active samples, it is a major player in shaping the cancer genome. In previous clinical studies, ~15% of pRCC patients responsed to cytotoxic chemo (REF) but we do not know who they are. Our APOBEC study and comparison to urothelial cancer help better understand the heterogeneity of the cancer nature. We were forming scientific hypotheses in the discussion section in hope to encourage further research ideas and interests. We completely understand the concern from the reviewer about the language and interpretation of the results. Therefore, in the revised manuscript, we rewrote this part to better distinguish actual results and our hypotheses.
Excerpt From Revised Manuscript	Given a statistically robust signal in our conservative algorithm, it is plausible that a small fraction of otherwise driver mutation absent type 2 pRCCs might share some etiologically and gnomically similarity with UC. Standard treatment for UC involves cytotoxic chemotherapy and radiation while RCC shows low response rate to cytotoxic therapy. Pending further research, this finding might lead to actionally clinical implications.

Shantao 2/24/2017 11:20 AM
Deleted: We want to emphasize that we are now doing explorations and forming hypotheses, trying to raise further research interests.

-- Ref3.4 – Significance of chromatin remolding defects --

Reviewer Comment	Papillary RCC with defects in chromatin remodeling genes show a higher mutation burden. This is interesting, but not too surprising as it is the case in other tumor types.
Author Response	We thank the reviewer for the comment. To our best knowledge, we are not aware of major systematic studies showing chromatin remolding (CR) defects are related with higher mutation burden in functionally important DHS regions. Most of the mutation burden studies focus on DNA repair genes. Besides, we showed CR genes mutations are not merely

	a refection of high mutation burden but associated directly with mutation landscape change. Out test statistics still stand when the mutation numbers in DHS regions are normalized by the total mutation counts.
	Last, we believe this key mutation landscape shift is an important source for tumor heterogeneity and intra-patients variations on genomic level. Understanding the heterogeneity and variations of pRCC is a key focus of our paper. Therefore, we believe our finding on CR defects inducing mutation landscape change is significant. We revised the manuscript to reflect the implication of our results.
Excerpt From Revised Manuscript	

-- Ref3.5 –Methylation analysis--

Reviewer Comment	That methylation influences mutation spectra is interesting and may be pursued, but it needs a more coherent story. Perhaps additional analyses on which mutation pathways are affected and any prognostic role?
Author Response	We thank the reviewer for the suggestions. 1. In the revised manuscript, we have added a downstream analysis of methylation-related mutations, emphasizing on the functional consequences of them. 2. During the revision, we realized the alternative splicing event observed in *MET* in the TCGA study is related to methylation. We showed the novel transcription isoform is due to L1 promoter activation, which is likely due to local hypomethylation. It also reflects global methylation dysfunction. Therefore, the novel *MET* isoform is associated with methylation cluster 1, which is further away from normal kidney tissues. 3. Methylation is the most important driving force of mutation spectra variation. It is a key factor in tumor heterogeneity. Our finding helps pin down the sources for tumor variations and understand tumor development.
Excerpt From Revised Manuscript	First we validated the TCGA identified methylation cluster 1 showed higher methylation lever than cluster 2 in all annotation regions (Figure S2, see Methods), prominently in CpG Islands (OR of sites being differentially hypermethylated: 1.29, 95%CI: 1.20-1.39, p<0.0001). We confirmed this association by showing samples from methylation cluster 1 had higher PC1 scores as well as higher C-to-T mutation counts and mutation percentages in CpGs (Figure 3C). This trend was further validated using a larger WXS dataset as well. Especially, the most hypermethylated group, CpG island methylation phenotype (CIMP), showed the greatest C-to-T in CpGs (Figure S2). As expected, C-to-T mutations in CpGs in group 1 showed higher but not statistically significant percentage overlapping with CpG islands compared with group 2 (1.8% versus 1.4%, p=0.14). Therefore, methylation status is the most prominent factor shaping the mutation spectra across patients. We further explored the functional impact of the excessive mutations driven by methylation. C-to-T mutations in CpGs were more likely to be in the coding region (OR=1.54, 95%CI: 1.27-1.85, p<0.0001) and

nonsynonymous (OR=1.47, 95%CI: 1.17-1.84, p<0.001). Yet, C-to-T mutations in CpGs did not show functional bias between two methylation groups in non-coding regions.

The TCGA study has identified a MET alternative translation isoform as a driver event (3). However, the etiology of this new isoform is unknown. We identified this isoform results from the usage of a cryptic promoter from an L1 element, likely due to a local loss of methylation (REF). This event was reported in several other cancer types (REF). To test its relationship with methylation, we found a closet probe (cg06985664, ~3kb downstream) on the Methylation array show marginally statistically significant (p=0.055, one-side rank-sum test). Additionally, as expected, this event is associated with methylation group 1 (odds ration (OR)= 4.54, p<0.041), indicating genome-wide methylation dysfunction. This association is stronger in type 2 pRCC and it shows a significant association with the C2b cluster (OR= 17.5, p<0.007)…

…Besides, in MET non-coding regions, we first find a cryptic promoter from a retrotransposon in the second intron initiates the alternative splicing event, which is classified as a driver event by the TCGA study (3). Methylation is a major source of silencing retrotransposon activities in human genome (REF). Indeed, we observed evidence for a local loss of methylation and global methylation dysregulation in samples expressing alternative isoforms. Therefore, we showed methylation change might drive pRCC growth through MET pathway.

-- Ref3.6 – Structural variation analysis --

Reviewer Comment	The structural variations were not explored in great detail. There were 343 SV events but were any recurrent? There were three cases carrying deletions in CDKN2 and 1 case with amplification in MET; otherwise, the structural variations appear as largely a negative result.
Author Response	We thank the referee for the feedback. In response to the concern raised by the referee, we reanalyze the SVs using a more refined approach in the revised manuscript.

We noticed the old alignment used an old version of aligner that does not support split read mapping and thus greatly compromised SV detection. Using high performance cluster, we were able to spend a great amount of CPU time to realign more than 100 billions of reads for higher quality mapping. Thanks to better performance with newly mapped reads and three additional WGS samples, our callset grew by ~30% to 424 SVs. Based on the number of SV event, 35 WGS samples clearly show distinct two types: chromosome unstable (>40 SVs) and stable (<10 SVs). The unstable type roughly matches the unstable type identified by array-based method in the TCGA study.

By intersecting the SVs with curated cancer consensus genes from COSMIC, we found two cases with SDHB deletion. These two samples show significantly lower SDHB expression level, which further confirmed our findings. SDHB is a mitochondrial protein in respiratory chain. Germline mutation is associated with paraganglioma and RCC. Some other interesting but sporadic SV cases include EGFR2 and HIF1A duplication, DNMT3A and STAG2 deletion |

	Moreover, our SVs generated with WGS confirmed 3/5 CDKN2A cases that were reported by the TCGA study based on SNP array. We are able to show that our SV result predicts CDKN2A expression better than array-based method. Three confirmed cases all show very low CDKN2A expression levels whereas the two unformed cases do not (Supplement figure XX). Last, we want to point out that our SV set from sequencing has much finer resolution than the original SNP-array based approach. Therefore, we are able to conduct analyses on breakpoints. Although MET is involved in a lot of amplification events and several samples are genomically unstable, surprisingly we do not find any breakpoint falls into MET and disrupt the gene. This further supports the oncogene role of MET in pRCC. We add our new SV analysis results in to the revised manuscript.
Excerpt From Revised Manuscript	We used DELLY (10) to perform structural variants (SVs) discovery from WGS reads information (see Methods and Table S3). The SV discovery approach has higher sensitivity and resolution than array-based methods, which were employed in the TCGA analysis. In the end we found 424 somatic SV events, includes deletions, duplications, inversions and translocations (Figure SXX). Based on the SV event number, samples clearly split into two types: genome unstable (>40 events) and genome stable (<10 events). First, by overlapping SVs with known pRCC related genes, we found two cases with deletion in SDHB. The median SDHB expression is only ~50% compared to cases without alternation (Figure SXX). We confirmed three cases carrying deletions affecting CDKN2A called by TCGA array-based methods but not the other two cases, possibly due to large-scale events (aneuploidy). Notably, three confirmed cases have significantly lower CDKN2A expression but not in the unconfirmed two cases. This suggests SV calling from WGS is accurate and predicts expression better. One sample, TCGA-B9-4116, which has extensive amplification of MET, showed multiple SVs of various classes hitting MET regions. However, surprisingly, we did not find SVs affecting MET except this one example. We postulate trisomy/polysomy 7 is the main mechanism of MET structural alteration rather than duplication in a smaller scale. Besides duplication, we did not expect to find deletion, inversion or translocation disrupting oncogene MET. These SVs are likely to cause loss-of-function rather than gain-of-function mutations. This is consistent with the putative role of MET as an oncogene, rather than a tumor suppressor. Last, we observed several interesting sporadic events, including duplications in EGFR and HIF1A duplication and deletions in DNMT3A and STAG2 (see SXX).