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 the concerns about the results significance of MET. Indeed, MET has been known to be the central driver for type I pRCC for decades. However, most of the analyses focus on 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, without evidence for missense mutations, alternative splicing and copy number amplification. Using a more integrated approach, we are able to provide hints for alternative mechanisms to MET dysfunction in type I pRCC. 1. We find more somatic mutations in an extended WXS set, further completing the MET mutation spectrum of the TCGA study. 2. We find excessive non-coding mutation at the 5' end of MET. Given the critical role MET plays in pRCC, we believe this mutation hotspot is possibly linked with pRCC molecular etiologies. 3. During our revision, we find the activation of a cryptic promoter in the second intron of MET causes the alternative isoform discovered in the original TCGA study. This event has been observed in several other cancers included CML and some GI (gastrointestinal tract) cancers. Our finding is the first one that provides an explanation for the alternative RNA isoform. 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 result with discussion in the revised manuscript.
Excerpt From Revised Manuscript	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, 95%CI: 1.07-19.34, 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, 95%CI: 1.72-32.6, 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 statistical test power is affected by the small cohort size and the pathological heterogeneity of the samples. However, our analysis in non-coding regions still provides some meaningful insights of pRCC. 1. In our revision process, we reviewed the WGS samples and added three more WGS samples into our cohort, reading a final size of 35. 2. The non-coding mutation hot spots indeed carry excessive 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 we discover show extremely high mutation rate in our cohort. The hotspots span from 7 to 50kb, with 6 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. 3. As the reviewer pointed out, small sample size and the heterogeneity nature of the cohort greatly affect our statistical power. We were not able to provide much supporting evidence for the functional effects of these mutation hotspots. However, all three functional regions are biologically linked with pRCC. Besides, non-coding regions are largely overlooked in the previous studies of pRCC. We hope our analyses will spark interests and encourage researcher to further explore the possible biological impacts of these events.
Excerpt From Revised Manuscript	

-- 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. Therefore, it was overlooked in previous studies of pRCC. There are several studies supporting the role of NEAT1 in a few

	other cancers. Moreover, during the revision process, we learn some preliminary results from other research groups. In the currently ongoing PCAGW study (PanCancer Analysis of Whole Genomes), researchers are powered by the unprecedentedly large whole genome sample size and highlight NEAT1 (along with its co-expression partner MALAT1) as a newly discovered cancer driver (unpublished data, personal correspondence). In our study, we have shown it is linked with higher expression of NEAT1, presumably due to the dysfunction of gene regulation region, and worse survival of patients. The referee raised an interesting point about expression signature. Unfortunately due to small sample size of whole genome sequencing, we could not find statistically significant classification or transcriptomic signature studies. NEAT1 mutations seem to be associated with RNAseq cluster 3 but Fisher exact test does not support a statistical significance ($p>0.05$). Although lacking WGS data to find genomic alteration, we found NEAT1 is overexpressed in about 6% of the TCGA KIRC cohort. And NEAT1 higher expression is significantly associated with shorter overall survival time (OS). NEAT1 also show high co-expression with MALAT1, which is another noticeable lncRNA in cancer.
Excerpt From Revised Manuscript	However, without mutation status, NEAT1 expression level is not significantly linked with pRCC survival. Nonetheless, NEAT1 is overexpressed in about 6% 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. 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. In the revised manuscript we add discussions to better explain the significance of this part of the study. Mutation spectra elucidate the mutation processes in cancer. In our study, we identify several factors (methylation, APOBEC, chromatin remodeling defects etc.) play vibrant roles in tumorigenesis. This helps to better characterize and understand pRCC in terms of mutagenesis, tumor evolution, and molecular etiologies.

	Mutation burden has important predictive value on immune therapy response. In the era of great advancing of immune therapy and given the fact that IL-2 was the mainstay of RCC systematic treatment for many years, we feel research on mutation landscape in pRCC has the potential to facilitate clinical decisions.
Excerpt From Revised Manuscript	Not too sure what to do with this

-- 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. In the revision, we build… (running on HPC)
Excerpt From Revised Manuscript	

-- Ref1.5 – 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) "Potentially, patients with rs11762213 might also benefit from MET inhibitors Should we fight back on this? Stating "*Type II patients carrying rs11762213 only constitute a small subset of the patients. Thus clinical trials were not able to rule out MET inhibitor might be effective in this subset. One following-up study could be stratifying the cohort based on rs11762213 genotype and reanalyze the data*" I think Brian has some ongoing MET inhibitor trials to support this. Or we just turn down the language? c) We expanded the section of NEAT1, see REF1.3.

Excerpt From Revised Manuscript	b)

-- 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 made a good point. A recent publication about rs11762213 by AA Hakimi et al. looked at MET expression patterns in ccRCC. They did not find statistically significant change in MET in ccRCC associated with this SNP. Also this SNP is not in strong linkage disequilibrium with other SNPs of interest in RCCs. MET, on the other side, is strongly linked with RCCs. Since this is a germline SNP, it may affect the 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. However, this could be due to low statistical power. ~~In the revised manuscript, we better elaborated the current research status of rs11762213.~~
Excerpt From Revised Manuscript	

-- 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 agree that, with chromatin remodeling dysfunction, the DHS regions are likely to change in pRCC tumors. DHS regions called from a normal kidney cell line represent the open chromatin regions under

	normal, physiological condition. The observed the mutation burden shift could be a result of DHS regions change. In fact, we believe this is the most plausible explanation for this mutation landscape shift since both DNA replication and repair are affected by DNA accessibility. DHS regions are enriched with function regions of genome. A higher mutation burden in DHS regions might be deleterious for tumor. Recent studies have shown patients with higher and impactful mutation burden response better to immunotherapy. Thus this shift of mutation landscape may have clinical implications.
Excerpt From Revised Manuscript	

-- 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.

NO TREAT.

Author Response	We totally agree with the reviewer that there is a long path to translate scientific discoveries in the lab into clinical care improvement. 1. The two previous studies about rs11762213 were done on a mixed RCC cohort and an entire ccRCC respectively. The mixed cohort was mostly made up by 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. In this manuscript, we proved that rs11762213 also has predictive value in type 2 pRCC outcome for the first time. 2. p-value indicates the chances that the null hypothesis is true. It is certainly impacted by the magnificence 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 that we should rewrite on this part. Thus we revised the SNPs discussion in the manuscript.
Excerpt From Revised Manuscript	

-- 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. Recurrent mutations in NEAT1 are indeed intriguing. NEAT1 is a non-coding RNA thus will be missed by whole exome sequence. Therefore, it is overlooked in previous studies of pRCC. There are several studies supporting the role of NEAT1 in several cancers. In our study, we have shown it is linked with higher expression of NEAT1, presumably due to the dysfunction of gene regulation region, and worse survival of patients. There are several studies supporting the role of NEAT1 in a few other cancers. Moreover, during the revision process, we learn some preliminary results from other research groups. In the currently ongoing PCAGW study (PanCancer Analysis of Whole Genomes), researchers are powered by the unprecedentedly large whole genome sample size and highlight NEAT1 (along

	with its co-expression partner MALAT1) as a newly discovered cancer driver (unpublished data, personal correspondence). In our study, we have shown it is linked with higher expression of NEAT1, presumably due to the dysfunction of gene regulation region, and worse survival of patients. Although lacking WGS data to find genomic alteration, we found NEAT1 is overexpressed in about 6% of the TCGA KIRC cohort. And NEAT1 higher expression is significantly associated with shorter overall survival time (OS). NEAT1 also show high co-expression with MALAT1, which is another noticeable lncRNA in cancer.
Excerpt From Revised Manuscript	"However, without mutation status, NEAT1 expression level is not significantly linked with pRCC survival. Nonetheless, NEAT1 is overexpressed in about 6% 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. 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. 1. APOBEC signature is intriguing in pRCC because it has been observed (in higher prevalence) in chRCC, not in ccRCC. Now for the first time we found evidence of it in pRCC. APOBEC activities are newly recognized in cancer development and could be linked with several cancer risk factors, including genetic predispositions and viral infection. 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. 2. In previous clinical studies, ~15% of pRCC patients response to cytotoxic chemo (REF) but we do not know who they are. Our APOBEC study and comparison to urothelial cancer are making efforts to better understand the heterogeneity of the cancer nature. (should we say this?) 3. We were forming scientific hypotheses here 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 elaborate our results more clearly.
Excerpt From Revised Manuscript	

-- 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	1. 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 proved 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. 2. Mutation spectra elucidate the mutation processes in cancer. In our study, we identify several factors (methylation, APOBEC, chromatin remodeling defects etc.) play vibrant roles in tumorigenesis. This helps to better characterize and understand pRCC in terms of mutagenesis, tumor evolution, and molecular etiologies. Last, mutation burden has important predictive value on immune therapy response.
Excerpt From Revised Manuscript	Similar to Ref1.4

-- 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. 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.
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). 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 tried to explore 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 nor in non-coding regions (Figure SXX). 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, 95%CI: 1.07-19.34, 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, 95%CI: 1.72-32.6, p<0.007).

-- 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	In the revised manuscript we reanalyze the SVs using a more sensitive approach. First, we recognized that the original BAM files were made by a very old version of BWA that does not support split-read mapping in alignment. Split-read are vital to SV detection. Therefore, we extracted all the reads from the BAMs, paired them and performed remapping. Then we applied LUMPY, a probabilistic SV caller based discordant read pairs and split reads to call the SVs. To evaluate the functional impacts of the somatic SVs, we used SVScore to prioritize and evaluate the SVs.
Excerpt From Revised Manuscript	