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 analyses focus on coding region only. 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. Our study is the first one that looks into the non-coding regions of pRCC. We find excessive non-coding mutations at the promoter and regulatory regions of MET. Given the critical role MET plays in pRCC, 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 RNA 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	The TCGA study identified a MET alternative translation isoform as a driver event (3). However, the etiology of this new isoform was unknown. We found this isoform results from the usage of a cryptic promoter from an L1 element, likely due to local loss of methylation (REF). This event was reported in several other cancer types (REF). To test its relationship with methylation, we found the 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

NC IS ACTIVE & INFIZED

Shantao 1/26/2017 10:00 PM
Deleted: the

Shantao 1/26/2017 8:03 PM
Deleted: the

Shantao 1/26/2017 10:00 PM
Deleted: for

Shantao 1/28/2017 7:39 PM
Deleted: Moreover, the

Shantao 1/26/2017 8:02 PM
Deleted: We find more somatic mutations in an extended WXS set, further completing the MET mutation spectrum of the TCGA study.

Shantao 1/26/2017 8:03 PM
Deleted: 5' end

Shantao 1/26/2017 10:02 PM
Deleted: discovered

Shantao 1/26/2017 10:03 PM
Deleted: Our finding is the first one that provides

Shantao 1/26/2017 10:04 PM
Deleted: result with discussion

Shantao 1/28/2017 7:50 PM
Deleted: has

Shantao 1/28/2017 7:50 PM
Deleted: is

Shantao 1/28/2017 7:49 PM
Deleted: identified

Shantao 1/28/2017 7:50 PM
Deleted: a

Shantao 1/28/2017 7:50 PM
Deleted: a

	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. However, we start with finding noncoding mutations with high impact. We are able to show a significant amount of samples carry impactful mutations in noncoding regions and conduct some coarse recurrence tests. As the referee points out, we are limited by low statistical power to do further refined analyses. Nonetheless, our analysis is the first exploration of pRCC non-coding regions and provides meaningful insights of 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. All three functional regions are biologically linked 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.
Excerpt From Revised Manuscript	

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

Reviewer	This reviewer was very intrigued by the NEAT1 finding, which deserves more

Shantao 1/28/2017 7:52 PM
Deleted: by the small cohort size and the pathological heterogeneity of the samples.

Shantao 1/26/2017 10:05 PM
Deleted: in

Shantao 1/26/2017 8:14 PM
Deleted: still

Shantao 1/26/2017 8:14 PM
Deleted: some

Shantao 1/26/2017 8:05 PM
Deleted:

Shantao 1/26/2017 8:05 PM
Deleted: <#>In our revision process, we reviewed the WGS samples and added three more WGS samples into our cohort, reading a final size of 35.

Shantao 1/26/2017 10:06 PM
Deleted: s

Shantao 1/28/2017 8:10 PM
Deleted: we discover show

Shantao 1/28/2017 8:10 PM
Deleted: <

Shantao 1/28/2017 8:14 PM
Deleted: 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, a

Shantao 1/28/2017 8:16 PM
Deleted: Besides

Shantao 1/28/2017 8:15 PM
Deleted:

Shantao 1/28/2017 8:15 PM
Deleted: W

Comment	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 about 6% of the TCGA ccRCC cohort. NEAT1 higher expression is significantly associated with shorter overall survival time (OS). NEAT1 is tightly co-expressed with MALAT1, 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 small sample size. NEAT1 expression pattern is… Additionally, we are a part of the currently ongoing PCAWG study (PanCancer Analysis of Whole Genomes). During revision, we quickly looked at the NEAT1 mutation stats in the PCAWG RCC dataset. 21/144(14.58%) of the samples carry mutations in NEAT1, a frequency agrees with the one from our cohort. Confirmed hypermutation in a larger, high-quality dataset further supports our results. Unfortunately, we are not able to publish results based on PCAWG data at this moment.
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. 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.

Shantao 1/26/2017 8:15 PM
Deleted: Therefore, it

Shantao 1/26/2017 8:17 PM
Deleted: . ... [1]

Shantao 1/26/2017 8:17 PM
Deleted: .

Shantao 1/26/2017 8:26 PM
Deleted: Moreover

Shantao 1/26/2017 8:22 PM
Deleted: during the revision process, we learn some preliminary results from other research groups. In

Shantao 1/26/2017 8:15 PM
Deleted: G

Shantao 1/26/2017 8:23 PM
Deleted: ,

Shantao 1/26/2017 8:23 PM
Deleted:

Shantao 1/26/2017 8:23 PM
Deleted: researchers

Shantao 1/26/2017 10:10 PM
Formatted: Not Highlight

Shantao 1/26/2017 8:25 PM
Deleted: 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. ... [2]

	Mutation spectra elucidate diversified mutation processes in pRCC. In our study, we identify several factors (methylation, APOBEC, chromatin remodeling defects etc.) play vi~~ta~~l 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 advancing of immune therapy, we feel research on mutation landscape in pRCC has the potential to facilitate clinical decisions. In the revised manuscript we add discussions to better explain the significance of this part of the study.
Excerpt From Revised Manuscript	

Shantao 1/26/2017 8:40 PM
Deleted: . ... [3]

Shantao 1/26/2017 8:38 PM
Deleted: 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*

Shantao 1/26/2017 8:40 PM
Formatted: Highlight

	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)

Shantao 1/26/2017 8:45 PM
Deleted: ,

Shantao 1/26/2017 8:45 PM
Deleted: s

-- 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. studies ~~the~~ 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. Following the suggestion of referee, we explored the genes within 50kb away from MET … 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 elaborate the current research status of rs11762213 and incorporate the discussions above.
Excerpt From Revised	

Shantao 1/26/2017 8:41 PM
Deleted: made a

Shantao 1/26/2017 8:41 PM
Deleted: good point

Shantao 1/26/2017 8:45 PM
Deleted: looked at MET expression patterns

Shantao 1/26/2017 8:46 PM
Deleted: in MET in ccRCC

Shantao 1/26/2017 10:13 PM
Deleted: MET, on the other side, is strongly linked with RCCs.

Shantao 1/26/2017 8:48 PM
Formatted: Not Strikethrough

Shantao 1/26/2017 8:49 PM
Deleted: d

Shantao 1/26/2017 8:48 PM
Deleted: .

Shantao 1/26/2017 8:48 PM
Formatted: Not Strikethrough

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	[STL2MG: What's the EncodeCA results on DHS when comparing cancer/normal cell lines? Also I am not very sure whether we should fight back on this. If we say DHS region shift only adds second-order effect, then how do we rationalize mutation rate change? Chromatin remodelers participating in DNA repair? Then why particularly high mutation rate in open chromatin region? I feel we could just step back and agree with the referee. Then it is just a language game: we call DHS as "open chromatin regions in normal state". We will then have a convincing rational on the mechanism and still show the results are still meaningful.] 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 slightly in pRCC tumors. This change could partially explain the observed the mutation burden shift, but more likely to be a second-order effect. DHS regions are enriched with function regions of 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 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.

Shantao 1/26/2017 9:02 PM
Formatted: Highlight

Shantao 1/26/2017 9:02 PM
Formatted: Highlight

Shantao 1/26/2017 9:02 PM
Formatted: Highlight

Shantao 1/26/2017 8:51 PM
Deleted: w

Shantao 1/26/2017 8:49 PM
Deleted: change

Shantao 1/26/2017 8:51 PM
Deleted: DHS regions called from a normal kidney cell line represent the open chromatin regions under normal, physiological condition. The

Shantao 1/26/2017 8:55 PM
Deleted:

Shantao 1/26/2017 8:55 PM
Deleted: could be a result of DHS regions change

Shantao 1/26/2017 8:54 PM
Deleted: .

Shantao 1/26/2017 8:54 PM
Deleted: 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.

Shantao 1/26/2017 9:03 PM
Deleted: A

Shantao 1/26/2017 10:15 PM
Deleted:

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 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. In this manuscript, for the first time, we find that rs11762213 has predictive value in type 2 pRCC outcome. 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.

Shantao 1/26/2017 10:17 PM
Formatted Table

Shantao 1/26/2017 9:05 PM
Deleted: improvement

Shantao 1/26/2017 9:06 PM
Deleted:

Shantao 1/26/2017 9:11 PM
Formatted: Font:Italic

Shantao 1/28/2017 8:53 PM
Deleted: n entire

Shantao 1/28/2017 8:54 PM
Deleted: made up by

Shantao 1/26/2017 9:04 PM
Deleted: proved

Shantao 1/26/2017 9:04 PM
Deleted: also

Shantao 1/28/2017 8:54 PM
Deleted: for the first time.

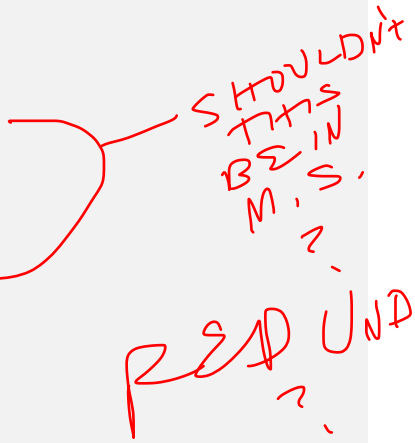
	In our case, the p-value is largely bounded by the small sample size. A “marginal” p-value does not necessarily mean the effect of the SNP on prognosis is small. In the revised manuscript, we calculated the odds ratio to better reflect the effect of rs11762213. 3. We were forming hypotheses and speculating 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	

Shantao 1/26/2017 9:12 PM
Formatted: No bullets or numbering

Shantao 1/26/2017 9:12 PM
Deleted: on

-- 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. The effect of NEAT1 is not minor. Patients generally have good prognosis in our pRCC cohort, thus affects the power of our survival analysis. However, in the revision, we looked at the TCGA ccRCC cohort. Although lacking WGS data to find genomic alteration, we found NEAT1 is overexpressed in about 6% of the cohort. NEAT1 higher expression is significantly associated with shorter overall survival time (OS). NEAT1 is tightly co-expressed with MALAT1, which is another noticeable lncRNA in cancer. Additionally, we are a part of the currently ongoing PCAWG study (PanCancer Analysis of Whole Genomes). During revision, we quickly looked at the NEAT1 mutation stats in the PCAWG RCC dataset. 21/144(14.58%) of the samples carry mutations in NEAT1, a frequency agrees with the one from our cohort. Confirmed hypermutation in a larger, high-quality dataset further



	supports our results. Unfortunately, we are not able to publish results based on PCAWG data at this moment. In the revised manuscript, we include the new analyses we have done on NEAT1 to support its role in pRCC.
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)."

Shantao 1/26/2017 9:21 PM
Deleted: 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. ... [4]

-- 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 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. We want to emphasize that we are now doing explorations and forming hypotheses, trying to raise further research interests. 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 better distinguish actual results and our hypotheses.

Shantao 1/26/2017 9:49 PM
Deleted: ... [5]

Shantao 1/28/2017 9:15 PM
Formatted: Not Highlight

Shantao 1/26/2017 9:51 PM
Formatted: Indent: Left: 0.5", No bullets or numbering

Shantao 1/26/2017 9:47 PM
Deleted: . (should we say this?)

Shantao 1/26/2017 10:19 PM
Deleted: elaborate

Shantao 1/26/2017 10:19 PM
Deleted: our

Shantao 1/26/2017 10:20 PM
Deleted: more clearly

GOOD STAT SUPP.

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	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.
Excerpt From Revised Manuscript	

Shantao 1/26/2017 9:59 PM
Formatted: No bullets or numbering

Shantao 1/28/2017 9:00 PM
Deleted: proved

Shantao 1/26/2017 9:59 PM
Deleted: ... [6]

Shantao 1/26/2017 9:59 PM
Deleted: 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, 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.
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	We understand the concern raised by the referee. First 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 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 driver role of MET in pRCC. [STL2MG: maybe we want to say this new SV analysis requires a lot of computation and work? So we should get at least an A+ for "effort grade"?] Also, in the revised manuscript, we reanalyze the SVs using a more refined approach. Using high performance cluster, we are able to spend a giant amount of CPU times to realign all the reads for higher quality mapping. We found… We update the manuscript to include the new SV analysis.
Excerpt From Revised Manuscript	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.

Shantao 1/26/2017 9:38 PM
Deleted: I

Shantao 1/28/2017 9:01 PM
Deleted:

Shantao 1/26/2017 11:01 PM
Deleted: sensitive

Shantao 1/26/2017 9:38 PM
Deleted: . 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.

Shantao 1/26/2017 9:38 PM
Deleted: To evaluate the functional impacts of the somatic SVs, we used SVScore to prioritize and evaluate the SVs.