Author’s response to reviews

Title: Predominance of Triple Wild-Type and IGF2R Mutations in Mucosal Melanomas

Authors:

Yuuki Iida (yiida-tky@umin.ac.jp)
Matthew Salomon (matthew.salomon@providence.org)
Keisuke Hata (qyp11547@gmail.com)
Kevin Tran (Kevin.Tran@providence.org)
Shuichi Ohe (shulohe@gmail.com)
Chester Griffiths (chestermd@aol.com)
Sandy Hsu (hopefulsmile@gmail.com)
Nellie Nelson (newilcox11@gmail.com)
Dave Hoon (hoon@jwci.org)

Version: 1 Date: 08 Sep 2018

Author’s response to reviews:

Please find our point by point responses to the reviewer’s comments directly below each comment in blue text.

Reviewer 1:

TERMS
In several locations the authors use the terms "mt"; to my knowledge "mutation" is more correct.
- We have changed “mt” to “mutation” throughout the manuscript.

The authors use the terms "mt burden"(lines 10,15 on pg 9); to my knowledge "Tumor mutation burden" (TMB) is more correct
- We have changed the term “mt burden” to “tumor mutation burden” on page 9 as well as throughout the manuscript.

**IHC**

It might be important to increase the number of IHC tests in order to verify whether there are differences of pd-l1 expression between MM and MC as well as between the different anatomical sites. Correlation between pd-l1 expression and BRAF, NRAS, or c-KIT mutational status should be also assessed. In other words, IHC data are not sufficient to make any significant correlation.

- We agree that it would be ideal to have IHC data for each of the samples that were sequenced in our study, unfortunately all the available pathology blocks were used in the initial IHC assays and we do not have any remaining blocks left that can be used for addition IHC assays. However, we do believe that the observation we made with respect to mutation tumor burden and PD-L1 expression is robust given the original sample size. To further illustrate the strength of our original result, we performed 100K permutations of our data set to generate a distribution of p-values. Below is a figure of the distribution of p-values from 100K tests with (the x-axis on a log scale). The horizontal red dotted line indicates the p-value of the original test. For this it is clear that the chances of observing a lower p-value are very low and the bulk of the distribution lies far to the right of the red line. Again, we believe that this further supports our initial conclusions.

We looked at the number of samples that had available IHC data and nonsynonymous mutations in BRAF, NRAS, and c-KIT. However, the number of samples with both IHC and a mutation in one of the three genes was very low. For BRAF mutated samples there were only 1 MM and 11 CM samples and there was even fewer samples for NRAS mutated, 2 MM and 7 CM, and c-KIT mutated, 1 MM and 0 CM. Therefore it is difficult to make any meaningful comparisons.

**NGS**

To determine the accuracy of the sequencing system, it will be necessary to provide as much details as possible when describing the results of non-synonymous mutations about:

-Phred score. Q scores are used to measure base calling accuracy, one of the most common metrics for assessing sequencing data quality.
We agree that considering only high-quality base calls (and alignments) in our mutation detection pipeline is very important. To insure this, we processed our raw fastq files through the full GATK/picard pipeline (please see methods section of the manuscript) including performing base quality score recalibration (BQSR) to the original Illumina base quality scores. This has been shown previously to produce more accurate quality scores that the original Illumina methods. Additionally, we examined the base quality scores used in each somatic mutation call that we included in our final call set. This information is contained within the BQ tag in the VCF files generated by MuTect. Below we show that 1) the vast majority of bases used for mutation calling had high base quality scores > 30 on the Phred scale and 2) that there was no significant difference between MM and CM sample types (p > 0.05). Therefore, we believe that these mutation calls were generated from high quality data.

-MAF is widely used in population genetics studies because it provides information to differentiate between common and rare variants in the population.

- In the manuscript we describe a set of filters that were applied to the raw variant calls prior to the downstream analysis results that are presented in the manuscript. Among the filters that were included was one that removed any variant that had a minor allele frequency of > 0.01 in any of the following population-based data sets; the 1000 Genomes Project, ESP6500, ExAC, and CG46 databases. Therefore, all the variants included in the final analyses were not observed to have a frequency > 0.01 in any of the above populations. Below is a plot of the maximum allele frequency (if previously observed in the above databases) to better illustrate that no mutational site included in our analysis had an observed allele frequency above the threshold of 0.01 or 0.05, both of which are common cutoff points used to define “rare” alleles in populations.

Reviewer 2:

Yuuki Iida and colleagues present the molecular study of Mucosal Melanomas (MM) by next-generation sequencing method in order to investigate its mutation profile. MMs are rare and aggressive non-cutaneous melanomas, non-related to sun exposure. They have already been studied together with other melanoma subtypes; however, studies in large MM groups are less frequent. Thus, this article adds new knowledge of MMs.
1) Page 5: DNA extraction

No information is given about the percentage of tumour cells into the extraction area. Even if most melanomas are cancers with high % of cancer cells, a large proportion of samples may have <80% of them. Furthermore, melanoma is a highly aggressive and heterogeneous tumour. The choice of tumour cell rich areas could give more relevant results and identify mutations with low frequency.

- In general, mucosal melanoma cells have fewer immune infiltrates than cutaneous melanomas thus greater purity. For tumor purity, we assessed frozen tissues that were dissected by a surgical pathologist from the original tumor surgery whereby, a representative tissue was made into a FFPE tumor block. Majority (> 90%) of the cells in the frozen tissues were melanoma cells. For FFPE tumor block analysis, we assessed H&E stained slides and performed micro-dissection of melanoma tumor cells. We have added this information to the revised manuscript. Please see page 5, line 21 to page 6, line 3.

2) Page 6: variant calling and data analysis

Variants with "mt allele frequency >10%, and coverage >20x were used for further analysis". 20x is rather a mild criteria for the selection of new variants. NGS could miss hot spot mutations even in frozen tissue with coverage as low as x50. Higher coverage should be expected in searching for new variants in less known cancers.

- We agree that adequate sequencing depth is crucial to accurately identify somatic mutations. In our filtering step we applied a sequencing depth threshold that required a minimum of 20X coverage at a particular site to include that site in our final call set. However, the vast majority of sites in our final call set had coverages much greater than 20X. The mean coverage of all mutation calls in our final call set was ~274X with a median coverage of 174X. The figure below illustrates the distribution of coverage for all mutations in our final call set. Therefore, we believe that we have, on average, sufficiently deep sequencing coverage to allow for accurate calls. We have added these coverage summaries to the methods section to help clarify this point. Please see page 7, line 3-4.

3) Page 8: Results

The authors state that the C>T substitutions were significantly less in MM (57.8%) than CM (64.8%) supporting less involvement of UV exposure in MM. This statement is rather confusing because we expect that the UV exposure has no effect on MM development. Some MM
specimens present signatures for tobacco exposure, and the authors conclude that smoking is a potential pathogenic factor in MM. Have the authors the information about the smoking history for these patients to confirm this suggestion? Other explanations are possible for the signature of tobacco exposure and the genital and anorectal localisation of MM cancers?

- The distinction that we were trying to make with this statement is that when you look at the six possible mutation types we observe that C→T mutations are less common in MM than in CM. However, what fraction of C→T mutations are the result of UV exposure as compared to the results of spontaneous deamination during aging is difficult to know. Therefore, when you take tri-nucleotide context into account (i.e. 96-way mutational spectrum) you see that the specific mutational spectrum that has been associated with UV damage is, as expected, higher in CM than in MM. We have re-written this section in an attempt to better clarify this point. Please see pages 8-9.

- Unfortunately, we do not have patient histories of smoking for these samples. This would be difficult as to how long the patient had the primary tumor for mucosal melanoma is much more difficult to know as compared to clearly visible cutaneous melanomas. We have updated the manuscript to reflect this. Please see page 9.

- We are not aware of any other mutational causes that would lead to the same signature.

4) DCC antibody should be tested to strengthen its role as a prognostic marker. How DCC mt modify protein expression and its role in apoptosis?

- We agree that it would be of value to include IHC results for DCC expression. However, we do not have access to any additional pathology blocks that we could use for IHC assays.

- This is not within the scope of the manuscript as the activation of pathway(s) of DCC mutations in melanoma is not well known.

5) Figure 1B abscissa legend is illegible.

- We have increased the text size of the x-axes and rearranged the panels in Figure 1 to increase readability.
6) Recently published NGS analyses in mucosal melanomas are not included and listed in the article.

- We apologies for this oversight and have now included several NGS melanoma references to the introduction of this revised manuscript. Please see page 4, lines 20-21.