Basic Reporting:

I mentioned in my last review that the authors consistently imply that there is an association between the two variants of interest and asthma when no statistical tests were performed.

1. In some ways this has been addressed. However, there are some notable exceptions.
2. Example:
    a. No association was found for allele or genotype frequency of the variant in the PACRG gene and asthma (p=1.0 and p=0.2).
    b. However, the authors still suggest that with a larger sample of individuals "the difference in frequency of the PACRG heterozygous genotype between asthmatic and non-asthmatic horses may **increase.**"
    c. I mentioned last time that the allele and/or genotype frequencies were likely not stable with the sample sizes reported. Why do the authors report a trend to a future experiment when the statistical test is not significant, and the number it is based on might not be stable?
    d. Sentences like this need to be removed or changed.

Experimental Design:

The purpose of this study was "to determine the utility of RNA-Seq to call gene sequence variants, and to identify sequence variants potentially associated with asthma". I will focus my review on these two aspects.

1. the accuracy of RNA-Seq data to determine DNA genotypes
2. the variant enrichment procedure and their association to asthma

We will focus on the second aspect in this section:

During the last revision I focused on the lack of statistical power and statistical tests to determine if SNPs or RNA expression was associated to asthma.

1. Now, the authors have performed statistical tests to determine the strength of association for (DNA) altered allele frequencies and altered genotype frequencies to asthma for two SNPs of interest (one in PACRG and one in RTTN).
2. Although these statistical tests were performed, it is difficult to trust the authors variant enrichment procedure to select these two variants.
3. Additionally, multiple testing correction was not considered as previously requested.
    a. The variant enrichment procedure (for variants of interest) starts by using 26,619 and 24,527 variants pre- and post- challenge in asthmatic horses and, 28,909 and 28,451 variants pre- and post- challenge in non-asthmatic horses.
    b. Of these variants they called "consensus sequence variants" within each group pre-asthma, post-asthma, pre-non-asthma, post-non-asthma, were called with SeqMule.
    c. Now **I assume** that they filtered the consensus variants per group to only consensus sequence variants unique to the asthmatic group pre- (2823) and post- (1788) challenge.
        i. This step is only mentioned in the Result, line 213: "The GATK variant calling and filtering workflow yielded 2823 and 1788 sequence variants present specifically in the asthmatic group pre- and post-challenge, respectively (Suppl. Figure 1)."
            1. Please add this to the methods (around line 165 or so) and clarify if these numbers were from *consensus* sequence

variants per group, when exactly this filter was applied, and if the variants needed to be unique to each group.

ii. **If this step was performed, this by-passed much of the multiple testing correction burden.**
iii. By calling consensus sequence variants (I'm still not sure if they are calling consensus nucleotide or consensus genotypes) they are calling variants that are the most frequent in each of the four groups. Then the authors only use **consensus** variants that are unique (I think unique because they made a Venn diagram, this is not explicitly stated) to asthmatic pre- and post- groups. This filter enriches for variants that will be associated with their desired alternative hypothesis.
iv. **The authors should not enrich for SNPs that will agree with the their alternative hypothesis prior to statistical analyses. Example: in a GWAS researches do not enrich for SNPs associated with their dependent variable prior to the analysis to reduce the multiple testing correction burden.**
v. **However, they could filter variants based on detrimental protein effects (the next step) prior to statistical analyses.**

d. Then I guess the authors use VEP and SIFT to identify variants that might affect protein structure. However, they do not state how many impactful variants are removed or kept by this filter. However, I assume that it is >=10 but <= 4611 (2823+1788).
e. After this step the authors enrich for variants present in all asthmatic individuals and not present in all non-asthmatic individuals which leaves them with 10 variants (again enriching for variants that will agree with the alternative hypothesis). Then by 'manual verification' 8 of the 10 variants are excluded and they are left with their two variants of interest in PACRG and RTTN.
    i. How are 80% of these variants excluded? Were these 8 SNPs genotyped incorrectly? This is disconcerting.

The variant enrichment procedure would greatly benefit from a flow chart depicting methods. Did I misinterpret these steps?

At a minimum the authors should have performed a statistical association with the asthma phenotype for some X number of SNPs, followed by multiple testing correction for these X comparisons, as mentioned in my previous revision.

This is important as the authors results conclude that there is a significant association between altered allele frequencies for the variant in RTTN p = 0.042 and asthma. However, after a Bonferonni correction for multiple comparisons of only two tests (required alpha = 0.05/2 = 0.025) this is no longer considered significant. Notably, in the current version of the paper two tests were performed for the two variants of interest.

Validity of Findings:

The authors only validate two of the RNA-seq SNPs with Sanger sequencing (in ~24 individuals). Since this is a main objective of the paper I expected more than two of the ~35,000 variants (I summed the numbers from the Venn diagram) confirmed. Therefore accuracy measurements of the RNAseq data to call DNA genotypes/SNPs is questionable.

Although it is good that the authors verified these two variants of interest with sanger sequencing.

I worry that the by enriching for variants that will tend to agree with their alternative hypothesis prior to testing improperly bypasses the multiple testing correction burden and is leading the authors, and potentially others to be more interested in these two variants than they should be. I am worried that the lack of statistical association to asthma is being disregarded to some extent, and therefore the relevance of asthma to these two variants is being inflated. If they state that there is a significant association to asthma for RTTN, then they should be held to the same statistical standards of other groups. Of course, it is possible that these variants are relevant to asthma, however, it seems that the statistics do not strongly support this conclusion when considering multiple testing.

Comments for the Author:

The genotypes were called for 26,619 and 24,527 variants pre- and post- challenge in asthmatic horses, and 28,909 and 28,451 variants pre- and post- challenge in non-asthmatic horses.

1. Line 165: The authors conclude that the "Difference in sequence variants before and after challenge were likely attributable to difference in gene expression, as previously discussed (39)."
2. Why would the difference in the number of SNPs pre- and post- challenge be discussed in their differential expression paper cited? I did not find any variants called in citation 39.
3. The difference in variants identified pre- and post- challenge are likely due to differences in alternative splicing, allele-specific expression, or simply a change in gene expression due to a time. The total quantity of RNA molecules expressed (as discussed in citation 39) is somewhat irrelevant when trying to justify the difference in variants called pre- and post- challenge. Rather, the total number of RNA molecules expressed can change due to allele specific expression etc.
4. I don't believe this was appropriately presented in the text. This should be discussed as a limitation.
5. **Another thought, the difference in the number of sequence variants shown in the Venn diagram is likely very misleading if I interpret the methods correctly. They called "consensus sequence variants per group" and therefore I assume the most frequent (genotype or base-pair) was selected per group. This does not mean that the sequence variants were absent from the other groups, but rather that it was not the most frequent. Therefore, by choosing the most frequent (consensus) variants per group they might be inflating the differences between groups. Is this correct?**

Line 172-173: what are existing sequence variants?

> "existing sequence variants were excluded"

What was the minimum depth required to call variants?

Please consider adding a dedicated limitations section.