A peer-reviewed version of this preprint was published in PeerJ on 11 October 2018.

View the peer-reviewed version (peerj.com/articles/5759), which is the preferred citable publication unless you specifically need to cite this preprint.

Tessier L, Côté O, Bienzle D. 2018. Sequence variant analysis of RNA sequences in severe equine asthma. PeerJ 6:e5759
https://doi.org/10.7717/peerj.5759

Sequence variant analysis of RNA sequences in severe equine asthma

Laurence Tessier [1, 2] , **Olivier Côté** [1, 3] , **Dorothee Bienzle** Corresp. [1]

[1] Department of Pathobiology, University of Guelph, Guelph, ON, Canada
[2] BenchSci, Toronto, ON, Canada
[3] BioAssay Works, Ijamsville, MD, United States

Corresponding Author: Dorothee Bienzle
Email address: dbienzle@uoguelph.ca

Background. Severe equine asthma is a chronic inflammatory disease of the lung in horses similar to low-Th2 late-onset asthma in humans. This study aimed to determine the utility of RNA-Seq to call gene sequence variants, and to identify sequence variants or potential relevance to the pathogenesis of asthma. **Methods.** RNA-Seq data were generated from endobronchial biopsies collected from 6 asthmatic and 7 non-asthmatic horses before and after challenge (26 samples total). Sequences were aligned to the equine genome with Spliced Transcripts Alignment to Reference software. Read preparation for sequence variant calling was performed with Picard tools and Genome Analysis Toolkit (GATK). Sequence variants were called and filtered using GATK and Ensembl Variant Effect Predictor (VEP) tools, and two RNA-Seq predicted sequence variants were investigated with both PCR and Sanger sequencing. Supplementary analysis of novel sequence variant selection with VEP was based on a score of <0.01 predicted with Sorting Intolerant From Tolerant (SIFT) software, missense nature, location within the protein coding sequence and presence in all asthmatic individuals. For select variants, effect on protein function was assessed with Polymorphism Phenotyping (PolyPhen) 2 and Screening for Non-Acceptable Polymorphism (SNAP) 2 software. Sequences were aligned and 3D protein structures predicted with Geneious software. Difference in allele frequency between the groups was assessed using a Pearson's Chi-squared test with Yates' continuity correction, and difference in genotype frequency was calculated using the Fisher's exact test for count data. **Results.** RNA-Seq variant calling and filtering correctly identified substitution variants in *PACRG* and *RTTN*. Sanger sequencing confirmed that the *PACRG* substitution was appropriately identified in all 26 samples while the *RTTN* substitution was identified correctly in 24 of 26 samples. These variants of uncertain significance had substitutions that were predicted to result in loss of function and to be non-neutral. Amino acid substitutions projected no change of hydrophobicity and

isoelectric point in PACRG, and a change in both for RTTN. For *PACRG,* no difference in allele frequency between the two groups was detected but a higher proportion of asthmatic horses had the altered *RTTN* allele compared to non-asthmatic animals.
Discussion. RNA-Seq was sensitive and specific for calling gene sequence variants in this disease model. Even moderate coverage (<10-20 cpm) yielded correct identification in 92% of samples, suggesting RNA-Seq may be suitable to detect sequence variants in low coverage samples. The impact of amino acid alterations in PACRG and RTTN proteins, and possible association of the sequence variants with asthma, is of uncertain significance, but their role in ciliary function may be of future interest.

Sequence variant analysis of RNA sequences in severe equine asthma

Laurence Tessier[1,#], Olivier Côté[1,*], Dorothee Bienzle[1,**]

[1]Department of Pathobiology, University of Guelph, 50 Stone Road East, Guelph, Ontario N1G 2W1, Canada

[#]Current address: BenchSci, 559 College St., Suite 201, Toronto, ON M6G 1A9

[*]Current address: BioAssay Works, 10075 Tyler Place, Suite 18, Ijamsville, MD 21754, USA

[**]Corresponding author: *dbienzle@uoguelph.ca*

ABSTRACT

Background. Severe equine asthma is a chronic inflammatory disease of the lung in horses similar to low-Th2 late-onset asthma in humans. This study aimed to determine the utility of RNA-Seq to call gene sequence variants, and to identify sequence variants or potential relevance to the pathogenesis of asthma.

Methods. RNA-Seq data were generated from endobronchial biopsies collected from 6 asthmatic and 7 non-asthmatic horses before and after challenge (26 samples total). Sequences were aligned to the equine genome with Spliced Transcripts Alignment to Reference software. Read preparation for sequence variant calling was performed with Picard tools and Genome Analysis Toolkit (GATK). Sequence variants were called and filtered using GATK and Ensembl Variant Effect Predictor (VEP) tools, and two RNA-Seq predicted sequence variants were investigated with both PCR and Sanger sequencing. Supplementary analysis of novel sequence variant selection with VEP was based on a score of <0.01 predicted with Sorting Intolerant From Tolerant (SIFT) software, missense nature, location within the protein coding sequence and presence in all asthmatic individuals. For select variants, effect on protein function was assessed with Polymorphism Phenotyping (PolyPhen) 2 and Screening for Non-Acceptable Polymorphism (SNAP) 2 software. Sequences were aligned and 3D protein structures predicted with Geneious software. Difference in allele frequency between the groups was assessed using a Pearson's Chi-squared test with Yates' continuity correction, and difference in genotype frequency was calculated using the Fisher's exact test for count data.

Results. RNA-Seq variant calling and filtering correctly identified substitution variants in *PACRG* and *RTTN*. Sanger sequencing confirmed that the *PACRG* substitution was appropriately identified in all 26 samples while the *RTTN* substitution was identified correctly in 24 of 26 samples. These variants of uncertain significance had substitutions that were predicted to result in loss of function and to be non-neutral. Amino acid substitutions projected no change of hydrophobicity and isoelectric point in PACRG, and a change in both for RTTN. For *PACRG,* no difference in allele frequency between the two groups was detected but a higher proportion of asthmatic horses had the altered *RTTN* allele compared to non-asthmatic animals.

Discussion. RNA-Seq was sensitive and specific for calling gene sequence variants in this disease model. Even moderate coverage (<10-20 cpm) yielded correct identification in 92% of samples, suggesting RNA-Seq may be suitable to detect sequence variants in low coverage

samples. The impact of amino acid alterations in PACRG and RTTN proteins, and possible association of the sequence variants with asthma, is of uncertain significance, but their role in ciliary function may be of future interest.

INTRODUCTION

Severe equine asthma (recurrent airway obstruction, heaves) is a chronic inflammatory lung disease caused by inhalation of environmental dust and microbial components [1]. Exacerbation of the disease triggers excessive mucus production, cough, neutrophilic airway inflammation, bronchial hyperreactivity, and bronchospasm. Recurrent exacerbations induce smooth muscle hyperplasia and hypertrophy, fibrosis and eventual irreversible airway remodeling [2-5].

Asthma in humans is recognized to be a heterogeneous disease that is classified considering genetic, molecular and clinical features [6,7]. Severe equine asthma is most similar to human severe, late-onset asthma characterized by absence of Th2 cytokines, and presence of neutrophilic inflammation and bronchial neutrophil chemokines [6,8]. Severely asthmatic horses do not have a hypersensitivity response [2] and efforts to associate equine asthma with a Th2 cytokine profile have yielded inconsistent or inconclusive results [9-17]. Mechanisms leading to the development of both severe equine asthma and late-onset low-Th2 severe asthma in humans remain largely undefined.

Interactions between genes and environmental factors have been recognized to contribute to development of equine asthma for many years [18]. Genetic factors likely reside in multiple gene sequence variants, and may be influenced by age and sex [18-20]. Several susceptibility sequence variants, haplotypes and regions have been associated with human asthma [21-34] but no specific markers have been identified in the late-onset low-Th2 sub-phenotype [6,8]. Similarly, genetic markers of equine severe asthma were identified in certain families, but were not significantly associated across different families and genetic backgrounds [20,35-38].

RNA-Seq is a promising approach for calling sequence variants concurrent with analysis of gene and allele-specific expression, alternative splicing, and pathways. In this study we investigated whether SNV detected by RNA-Seq were also present in Sanger-sequenced amplicons. We hypothesized that RNA-Seq would identify gene sequence variants with high accuracy.

METHODS

Animals and procedures

Animal and sample procedures were previously described [39]. Briefly, six asthmatic and seven non-asthmatic horses without signs of asthma exacerbation belonging to the institutional research

herd (mean ages of 15 and 12 years, respectively, p =0.352, unpaired t test) were placed indoor in a dust-free environment. After 24 hours, asthmatic horses were exposed to dusty hay until exacerbation (range 1 to 3 days, average 2.2 days), while non-asthmatic horses were exposed for 3 days. Before and after the dusty hay asthmatic challenge, physical examination, pulmonary function test, and bronchoalveolar lavage were performed, and endoscopic bronchial biopsies were collected from lung lobes contralateral between first and second samples. Samples from an additional four asthmatic and seven non-asthmatic horses were used for PCR-amplification of specific sequence variant regions and Sanger sequencing. All procedures were approved by the Institutional Animal Care Committee of the University of Guelph (protocol R10-031) and conducted in compliance with Canadian Council on Animal Care guidelines.

RNA-Seq sample preparation and sequence alignment

RNA extraction, preparation and sequencing procedures were as described previously [39]. In brief, total RNA was extracted from endobronchial biopsies (Qiagen, Toronto, ON) and tested for quality and concentration with the Bioanalyzer RNA Nanochip (Agilent, Mississauga, ON) and capillary electrophoresis. RNA-Seq library preparation (unstranded) and sequencing were performed using the Illumina TruSeq RNA sample preparation and appropriate sequencing protocols (Illumina, San Diego, CA) at The Centre for Applied Genomics (TCAG; Toronto, ON). Sequencing of 100-base paired-end reads was performed following the manufacturer's instructions on an Illumina HiSeq 2500 instrument.

FastQC software version 0.10.1 (bioinformatics.babraham.ac.uk/projects/fastqc/) was used to assess quality of raw reads, and alignment to the horse reference genome [40] (Ensembl v70) was performed with STAR version 2.4 [41]. Specifically, the STAR_pass2 alignment protocol was followed using the horse Ensembl version 70 GTF annotation file for first- and second-pass, and the junction SJ.tab file generated by STAR for the second-pass after non-canonical junctions were removed. Default settings were used except for: --runThreadN 8 --outFilterScoreMinOverLread 0.5 --outFilterMatchNminOverLread 0.5. Details and results for read alignment were previously described [39].

Sequence variant calling and filtering

Read processing, sequence variant calling and initial filtering were performed following the Genome Analysis ToolKit (GATK) best practice guide for variant calling on RNA-Seq, except for the Indel realignment step considering the pass-2 STAR alignment initially performed. Initial read processing was first performed with Picard tools version 1.114 (broadinstitute.github.io/picard/) to add read groups and mark duplicates. Split n' Trim as well as base recalibration were performed using the GATK software version 3.2.2 [42] and the *-T SplitNCigarReads*, *-rf ReassignOneMappingQuality*, *-RMQF 255*, *-RMQT 60* and *-U ALLOW_N_CIGAR_READS* options.

The GATK variant calling and filtering workflow yielded 2,823 and 1,788 sequence variants present in all horses of the asthmatic group pre- and post-challenge, respectively (Suppl. Figure 1). Sequence variants were subsequently called using the Haplotype Caller function in GATK with the same genome annotation file used in the read alignment phase and the following options: *-recoverDanglingHeads*, *-dontUseSoftClippedBases*, *-stand_call_conf 20.0* and *-stand_emit_conf 20.0* options. Resultant sequence variants were processed with the variant filtration function of GATK software and the following options to establish a confidence threshold of reported variants: *-window 35, -cluster 3, -filterName FS, -filter "FS > 30.0", -filterName QD* and *-filter "QD < 2.0"*. Sequence variants were analyzed individually in each of 26 samples (6 asthmatics and 7 non-asthmatics, before and after asthmatic challenge).

PCR

Primers for amplification of sequence variant regions from bronchial DNA were parkin co-regulated *(PACRG)* forward (5′-CTC TGA ACC TCC GAA ACC GAC-3′) and reverse (5′-CTC CTG GGA TAA CTC ACC ATT C-3′), and rotatin (*RTTN)* forward (5′-TCC TGA GTT GTA TCA AGA AGT G-3′) and reverse (5′-CCA GCC TGC AAT TCC TTT CT-3′). A Taq polymerase PCR kit (Invitrogen, Mississauga, ON) was used for PCR amplifications. Each reaction was performed in a 25 µL final volume, including 5 µL of 10X PCR buffer, 0.2 mM dNTPs, 2 mM $MgSO_4$, 0.3 µM of each primer, 2 U of Platinum Taq, and 5 µL (100 ng) of template DNA. PCR conditions for amplification were 3 min at 94 °C followed by 35 cycles of 94 °C for 45 s, 60 °C or 58 °C for 30 s for *PACRG* and *RTTN*, respectively, and 72 °C for 90 sec, followed by final elongation for 10 min at 72 °C. Twenty µL of each PCR product was separated by electrophoresis in a 1% agarose gel stained with SYBR Safe (Invitrogen). Amplicons of

appropriate size were cut out and DNA extracted and purified (QIAquick, Qiagen). Extracted and purified PCR products were Sanger sequenced (Laboratory Services Division, Guelph, ON).

RESULTS

Sequence variant calling and filtering

The mean of the total number of RNA-Seq reads for all samples was 36,252,701.08, and the mean of uniquely mapped number of reads was 33,127,466.35. The number of individual total reads, total mapped reads, uniquely mapped reads and multiple mapped reads is detailed elsewhere (39). The GATK workflow resulted in 2,823 and 1,788 sequence variants present specifically in the asthmatic group pre- and post-challenge, respectively (Suppl. Figure 1). Of these, 10 were missense substitution variants, coded for proteins and had SIFT scores <0.01. Substitution variants in *PACRG* (Fig. 1A) and *RTTN* (Fig. 1B) were detected at higher proportion in asthmatic compared to non-asthmatic horses. A missense G/A substitution was detected in the coding sequence of *PACRG* at position 265,643 (Ensembl sequence ENSECAG00000014308) /264,806 (NCBI accession number 100050378) (Fig. 1A). A missense T/A substitution was detected in the coding sequence of *RTTN* at position 27,190 (Ensembl sequence ENSECAG00000009711)/ 27,871 (NCBI accession number 100052029) (Fig. 1B).

Amino acid sequence alignment

In PACRG, the G/A substitution resulted in replacement of valine (V) for methionine (M) at position 182 (Fig. 2A). PACRG sequence alignment of wild type (WT) and altered proteins predicted changes from beta-strand to alpha-helix structure in the altered protein a few amino acids distant from the site of substitution (182) at positions 187 and 188 (Fig. 2B). Hydrophobicity and isoelectric point were expected to remain similar despite the substitution (Geneious). In the RTTN sequence alignment, T/A substitution resulted in replacement of arginine (R) with tryptophan (W) at position 1807 of the ENSECAT00000010304 protein isoform (Ensembl sequence, corresponding to position 1812 of isoform X1 in NCBI [XP_001493238]) in NCBI sequence) (Fig. 3A). Sequence alignment of WT and altered proteins indicated a change from alpha helix to beta strand structure near the site of substitution (bp 1,807) at position 1,816 (Fig. 3B). In addition, increased hydrophobicity and decreased pI were projected at the site of substitution (1,807) in the altered compared to the WT protein.

Expression of *PACRG* and *RTTN*

Counts per million (cpm) for *PACRG* ranged from ~30.7 to 66.3 (mean = 44.17) in asthmatic horses (Fig. 4A) and ~25.1 to 65.6 (mean = 39.31) in non-asthmatic horses (Fig. 4B), while expression of *RTTN* varied from ~11.5 to 24.9 (mean =17.61) and ~13.1 to 29.9 (mean = 21.46) in asthmatic (Fig. 4C) and non-asthmatic (Fig. 4D) horses, respectively.

Confirmation of RNA-Seq with DNA Sanger sequencing

The *PACRG* substitution variants identified by RNA-Seq were confirmed on DNA with Sanger sequencing of PCR amplicons in 10 asthmatic (Fig. 5A) and 14 non-asthmatic (Fig. 5B) horses. In the asthmatic group, four horses were heterozygous [A/G] and two were homozygous with altered alleles [A/A] (Fig. 5A). In the non-asthmatic group, one horse was heterozygous with alleles [A/G], three horses were homozygous-WT [G/G] and three horses were homozygous-altered [A/A] (Fig. 5B). DNA was not available to confirm the genotype of horse 1. Hence, all genotypes derived from RNA-Seq were identical to Sanger sequencing results.

For *RTTN,* results of RNA-Seq and Sanger sequencing were very similar (Fig. 6). Four asthmatic horses were heterozygous [A/T] and two were homozygous altered [A/A] (Fig. 6A). In the non-asthmatic group, horse 5 was heterozygous [A/T] and four horses were identified as homozygous WT [T/T]. However, horse 1 was identified as homozygous before and heterozygous after the asthmatic challenge, while horse 4 was identified as heterozygous before and homozygous after the asthmatic challenge. Based on Sanger sequencing the genotype of horse 4 was homozygous. DNA was not available to confirm the genotype of horse 1. In all asthmatic horses, genotypes were consistent between pre- and post-challenge and sequencing methods.

Sequence alignment

Sanger DNA sequences of *PACRG* from 10 asthmatic horses and 14 non-asthmatic horses (including those that were analyzed by RNA-Seq) were aligned. Among asthmatic horses, eight had the heterozygous [A/G], two had the homozygous altered [A/A] and none had the homozygous WT [G/G] genotype. Among non-asthmatic horses, six had the heterozygous [A/G], five had the homozygous altered [A/A] and three had the homozygous WT [G/G] genotype (Fig.

7A). *RTTN* alignment in asthmatic horses yielded five heterozygous [A/T], two homozygous altered [A/A], and three homozygous WT [T/T] genotypes. In non-asthmatic horses, four had the heterozygous [A/T], 10 had the homozygous-WT [T/T], and no horse had the homozygous altered [A/A] genotype (Fig. 7B).

DISCUSSION

The goal of this study was to assess the reliability of an adapted RNA-Seq sequence variant calling workflow compared to Sanger sequencing. Sequence variant calling using RNA-Seq reads is recent practice, and reliability of results is a function of sequencing platform, depth, quality, precision of read mapping, and appropriate sequence variant calling and filtering methods. The reliability of identifying gene sequence variants using RNA-Seq has been considered uncertain. In some reports RNA-Seq was considered useful for identifying gene variants [54,55] while in other reports differences between RNA and DNA sequences were relatively frequent [56-58].

In this study we applied a modification of GATK best practices for sequence variant calling with RNA-Seq, and verified the results with Sanger sequencing. In 24 of 26 samples substitution variants in *PACRG* and *RTTN* were identified by both methods, while two horses' genotypes were discordant by RNA-Seq with inconsistent genotypes before and after challenge. Sanger sequencing confirmed one of the discordant genotypes, while the other could not be further assessed.

Two candidate substitution variants in the *PACRG* and *RTTN* coding sequence were identified after stringent filtering. Presence of the substitution variants was confirmed with PCR and Sanger sequencing in 24 samples. Correlation between RNA-Seq and Sanger sequencing showed that for *PACRG* both alleles of the gene were properly identified in all horses and conditions by the modified GATK workflow. For *RTTN*, two of the samples were misidentified by the workflow with alleles inconsistently identified before and after challenge. Lower mean expression suggesting lower sequencing coverage for *RTTN* might have affected the likelihood of inaccurate sequence variant calling. Nonetheless, the vast majority of alleles were identified appropriately, suggesting that the workflow is suitable for sequence variant calling in RNA-Seq at gene coverage in the 10 to 20 cpm range.

For supplementary analysis, SIFT was initially applied, followed by PolyPhen2 and SNAP2, to predict sequence variant effects on protein function for both substitution variants. SIFT uses phylogenetic data [45-49], while PolyPhen2 uses structural information and multiple alignments [50] to predict whether or not a sequence variant may cause loss of function. The two methods often yield similar results, but limited specificity suggests that results should be interpreted with caution [59]. SNAP2, on the other hand, uses evolutionary, structural, solvent-access and annotation information, as well as data from available homologs to predict whether a sequence variant is likely to have an effect on protein function [51-53]. While these three approaches can yield different results [60,61], inferences regarding PACRG and RTTN amino acid substitutions were consistent. However, ultimately conclusions regarding the effect of substitution variants require stringent protein functional analysis, and results from this study should be considered preliminary.

The substitutions identified changed V182M (valine to methionine) and R1807W (arginine to tryptophan) in PACRG and RTTN, respectively. For PACRG, the V->M substitution minimally affected hydrophobicity and pI, while the R->W substitution in RTTN increased hydrophobicity and decreased pI. The substitution variants were considered to potentially cause loss of function and to have non-neutral effects (Suppl. Tables 1 and 2). *PACRG* is a gene conserved across species [62] that shares a bi-directional promoter with parkin (*PARK2*) [63]. PACRG is affiliated with axonemal doublet microtubules, and contributes to the signaling pathway that controls dynein-driven microtubule sliding [62,64-66]. A single nucleotide variant (SNV) in *PACRG* was strongly associated with an increased risk of developing childhood asthma following early-life tobacco smoke exposure [67].

For the RTTN substitution variant, tryptophan is an aromatic, non-polar and hydrophobic amino acid often buried in hydrophobic cores, while arginine is a polar and positively charged amino acid often found on outside chains [68]. *RTTN* is a cilium-associated protein [69] essential for assembly of centrosomes in non-motile and motile cilia [70]. Absence of RTTN, or presence of gene sequence variants that disrupt the interaction of RTTN with SCL/TAL1 interrupting locus (*STIL*), abrogate proper ciliary development and function [70], and recessive mutations in *RTTN* are linked to abnormal primary ciliary development in humans [71]. A change in the structural stability or binding affinity of the entire protein or the affected residue could impact ciliary structure and function. The R1807W substitution variant in the carboxy-

terminal region is not immediately proximal to the suggested centrosome-targeting and STIL-binding site [70] but could nevertheless result in defective centrioles and hence affect cilium structure and function.

Substitution variant sequence determination in 13 RNA-Seq and 11 additional samples showed that 80% of asthmatic animals were heterozygous and 20% were homozygous-altered for PACRG, and that no individual had the homozygous-WT genotype (G/G). Conversely, among non-asthmatic animals more than half were homozygous, whether WT or altered (5 [A/A] and 3 [G/G]). For *RTTN*, 20% of asthmatic horses were homozygous-altered (A/A), 30% were homozygous WT (T/T) and 50% were heterozygous. Among non-asthmatics, none was homozygous altered (A/A), while 71% of horses were homozygous WT (T/T) and 29% were heterozygous. Therefore, the substitution was present in 70% of asthmatic horses and in only 30% of non-asthmatic horses (heterozygous or homozygous altered). Albeit, the variants have been identified in only a small sample of asthmatic and non-asthmatic animals, and have to be considered as variants of unknown significance (VUS). A comprehensive genome-wide association study (GWAS) would be necessary to determine association between these VUS and asthma, and statistical analysis of potential associations would need to be performed prior to filtering of variants.

Pearson's Chi-squared test with Yates' continuity correction applied detected no difference in allele frequency for *PACRG*, or in genotype frequency for *RTTN* and *PACRG*, between asthmatic and non-asthmatic horses. A significantly higher frequency of the altered allele (A) in asthmatic compared to non-asthmatic horses was identified. For *PACRG*, although not significant, the *p*-value obtained after testing for differences in genotype frequency ($P = 0.213$) was lower than when testing for allele frequency ($P = 1$). This finding may be attributed to the higher proportion of asthmatic horses with a heterogeneous genotype (WT/alt for eight of ten horses) compared to non-asthmatics (WT/alt for six of 14 horses). However, changes in allele frequency and potential roles in the pathogenesis of asthma are of unknown significance due to the small sample size in this study. Notwithstanding, a significant difference in the frequency of the *PACRG* heterozygous genotype has been reported in pulmonary tuberculosis in humans [72]. A genome-wide interaction study also identified a *PACRG* SNP to be linked to an increased risk of developing childhood-onset asthma following early-life exposure to tobacco smoke [67]. SNPs in *PACRG* also contributed to susceptibility to tuberculosis [73].

For *RTTN*, the difference in allele frequency was encouraging and further analysis with a larger number of samples to assess association with asthma may be warranted. RTTN is a centrosome-associated protein first discovered for its role in axial rotation and left-right specification in the mouse embryo [69]. The R->W substitution altered the hydrophobicity and isoelectric point at position 1807, and R <=> W substitutions were predicted to be most problematic in the human genome [74]. In addition, R->W substitution is generally disfavored in all protein types [75]. Overrepresentation of mutated arginine was a prominent feature among disease-causing mutations in a range of conditions [76].

As landmarks in epithelial-environmental interaction, cilia are highly specialized cellular projections. Most vertebrate cells have a single non-motile ('primary') cilium that transduces signals from the environment or other cells, while motile cilia occur in multiples on specialized cells of the respiratory tract, oviduct and ventricles of the brain [77,78]. Motile cilia directionally propel cells or extracellular fluid through "metachronal wave" beating movements [79,80]. The ability of motile cilia to beat in a synchronized manner requires specialized proteins that are absent in non-motile primary cilia, but otherwise both types of cilia have similar internal architecture. The main part of the cilium is the axoneme, which is comprised of nine outer microtubule doublets, one central microtubule pair (in motile multiple cilia) and a multitude of affiliated proteins. Prominent among these are tektins that stabilize microtubules and regulate axoneme length [81], and protofilament ribbon proteins that are essential for sliding of adjacent microtubule doublets to generate ciliary movement [82]. Abnormalities in cilia are now appreciated as cause for the development of respiratory diseases, often through gene sequence variants associated with a loss of function affecting unique ciliary proteins [77]. Ultrastructural changes were previously reported in the ciliated epithelium of horses with severe asthma (formerly called chronic obstructive pulmonary disease [COPD]), and included loss of ciliated cells [83]. Factors that affect beating, synchronization or orientation of motile cilia result in accumulation of mucus in airways [77], which is a prominent feature of equine asthma. Furthermore, hedgehog (HH) signaling is strongly linked to ciliary function, and many components of the HH signaling pathway localize to cilia [84,85]. However, considering the relatively small number of individuals tested, allele frequencies identified in this manuscript, and their potential impact on ciliary function, remain to be confirmed on a larger scale.

PACRG may be linked to HH signaling in mice where patched1 (*PTCH1*) and *PACRG-PARK2* loci are thought to interact and regulate ciliary function in ependymal cells [86]. Interestingly, *PTCH1* is differentially expressed in asthmatic compared to non-asthmatic horses following challenge [39], linking *PACRG* and an asthmatic response to environmental agents with the HH pathway. The PACRG protein associates with protofilaments [87] of the ciliary axoneme [62,88,89], has a role in ciliary morphogenesis and function [65] and is directly involved in ciliary motility through control of dynein-driven microtubule sliding [64]. PACRG also has a variety of interacting partners such as microtubules, α- and β-tubulin and meiosis/spermiogenesis associated 1 (MEIG1) protein, heat shock protein (HSP) 70 and HSP 90 [66,90,91]. Impaired function or interaction of PACRG with its partners could weaken or impair ciliary stability and motility. The exact nature and function of methionine in protein structure remains incompletely understood, and substitutions involving methionine has been associated with several diseases [92]. Both valine and methionine are hydrophobic residues grouped among the least polar amino acids [93]. Methionine is a sulfur-containing amino acid that is among the most hydrophobic residues and also easily oxidized if exposed [94]. Although V->M substitutions are generally neutral, methionine's sulfur connected to a methyl group would make it less likely to interact with other proteins [75]. Methionine was overrepresented as a mutant residue in several mutations associated with decrease or loss of function [76], including the human androgen receptor [95]. Although the effect of a V->M substitution is unknown, any change in PACRG structure or binding affinity could impact ciliary function, and may be of great interest in the context of severe asthma.

CONCLUSIONS

Sequence variants can be confidently called with RNA-Seq, although the required minimal coverage remains to be clearly defined and may be variable. Single point substitution variants in *PACRG* and *RTTN* were detected in all asthmatic horses, and although there was no significant difference in allele and genotype proportions between the two groups, the altered allele in the RTTN gene was more prevalent in asthmatic compared to non-asthmatic horses. Functional cilia are crucial for lung health, and sequence variants resulting in impaired protein function are likely

to have a negative impact. The significance of the substitutions in *PACRG* and *RTTN* remains to be determined but they are of potential interest for future investigations.

ACKNOWLEDGEMENTS

The authors thank Laurent Viel, Mary Ellen Clark and Andrés Diaz-Méndez for assistance with animal handling, biopsy collection and respiratory function measurements.

REFERENCES

1. Couëtil LL, Cardwell JM, Gerber V, Lavoie J-P, Léguillette R, Richard EA. Inflammatory Airway Disease of Horses--Revised Consensus Statement. J Vet Intern Med. 2016;30:503-15.
2. Pirie RS. Recurrent airway obstruction: a review. Equine Vet J. 2014;46:276-88.
3. Martinez FD, Vercelli D. Asthma. Lancet. 2013;382:1360-72.
4. Vargas A, Roux-Dalvai F, Droit A, Lavoie J-P. Neutrophil-Derived Exosomes: A New Mechanism Contributing to Airway Smooth Muscle Remodeling. Am J Respir Cell Mol Biol. 2016;55:450-61.
5. Setlakwe EL, Lemos KR, Lavoie-Lamoureux A, Duguay J-D, Lavoie J-P. Airway collagen and elastic fiber content correlates with lung function in equine heaves. Am J Physiol Lung Cell Mol Physiol. 2014;307:L252-60.
6. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. Nat Med. 2012;18:716-25.
7. Skloot GS. Asthma phenotypes and endotypes: a personalized approach to treatment. Curr Opin Pulm Med. 2016;22:3-9.
8. Ilmarinen P, Tuomisto LE, Kankaanranta H. Phenotypes, Risk Factors, and Mechanisms of Adult-Onset Asthma. Mediators Inflamm. 2015;2015:514868.
9. Lavoie JP, Maghni K, Desnoyers M, Taha R, Martin JG, Hamid QA. Neutrophilic airway inflammation in horses with heaves is characterized by a Th2-type cytokine profile. Am J Respir Crit Care Med. 2001;164:1410-3.
10. Joubert P, Silversides DW, Lavoie JP. Equine neutrophils express mRNA for tumour necrosis factor-alpha, interleukin (IL)-1beta, IL-6, IL-8, macrophage-inflammatory-protein-2 but not for IL-4, IL-5 and interferon-gamma. Equine Vet J. 2001;33:730-3.

406 11. Klukowska-Rötzler J, Marti E, Lavoie J-P, Ainsworth DM, Gerber V, Zurbriggen A, Janda J.
407 Expression of thymic stromal lymphopoietin in equine recurrent airway obstruction. Vet
408 Immunol Immunopathol. 2012;146:46-52.
409 12. Ainsworth DM, Grünig G, Matychak MB, Young J, Wagner B, Erb HN, Antczak DF.
410 Recurrent airway obstruction (RAO) in horses is characterized by IFN-gamma and IL-8
411 production in bronchoalveolar lavage cells. Vet Immunol Immunopathol. 2003;96:83-91.
412 13. Debrue M, Hamilton E, Joubert P, Lajoie-Kadoch S, Lavoie J-P. Chronic exacerbation of
413 equine heaves is associated with an increased expression of interleukin-17 mRNA in
414 bronchoalveolar lavage cells. Vet Immunol Immunopathol. 2005;105:25-31.
415 14. Ainsworth DM, Matychak M, Reyner CL, Erb HN, Young JC. Effects of in vitro exposure to
416 hay dust on the gene expression of chemokines and cell-surface receptors in primary bronchial
417 epithelial cell cultures established from horses with chronic recurrent airway obstruction. Am J
418 Vet Res. 2009;70:365-72.
419 15. Ainsworth DM, Wagner B, Erb HN, Young JC, Retallick DE. Effects of in vitro exposure to
420 hay dust on expression of interleukin-17,-23,-8, and-1β and chemokine (CXC motif) ligand 2 by
421 pulmonary mononuclear cells isolated from horses chronically affected with recurrent airway
422 disease. Am J Vet Res. 2007;68:1361-9.
423 16. Lavoie-Lamoureux A, Moran K, Beauchamp G, Mauel S, Steinbach F, Lefebvre-Lavoie J,
424 Martin JG, Lavoie J-P. IL-4 activates equine neutrophils and induces a mixed inflammatory
425 cytokine expression profile with enhanced neutrophil chemotactic mediator release ex vivo. Am
426 J Physiol Lung Cell Mol Physiol. 2010;299:L472-82.
427 17. Padoan E, Ferraresso S, Pegolo S, Castagnaro M, Barnini C, Bargelloni L. Real time RT-
428 PCR analysis of inflammatory mediator expression in recurrent airway obstruction-affected
429 horses. Vet Immunol Immunopathol. 2013;156:190-9.
430 18. Marti E, Gerber H, Essich G, Oulehla J, Lazary S. The genetic basis of equine allergic
431 diseases 1. Chronic hypersensitivity bronchitis. Equine Vet J. 1991;23:457-60.
432 19. Couëtil LL, Ward MP. Analysis of risk factors for recurrent airway obstruction in North
433 American horses: 1,444 cases (1990-1999). J Am Vet Med Assoc. 2003;223:1645-50.
434 20. Ramseyer A, Gaillard C, Burger D, Straub R, Jost U, Boog C, Marti E, Gerber V. Effects of
435 genetic and environmental factors on chronic lower airway disease in horses. Journal of
436 veterinary internal medicine. 2007;21:149-56.

21. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, Depner M, von Berg A, Bufe A, Rietschel E, Heinzmann A, Simma B, Frischer T, Willis-Owen SAG, Wong KCC, Illig T, Vogelberg C, Weiland SK, von Mutius E, Abecasis GR, Farrall M, Gut IG, Lathrop GM, Cookson WOC. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature. 2007;448:470-3.

22. Choudhry S, Taub M, Mei R, Rodriguez-Santana J, Rodriguez-Cintron W, Shriver MD, Ziv E, Risch NJ, Burchard EG. Genome-wide screen for asthma in Puerto Ricans: evidence for association with 5q23 region. Hum Genet. 2008;123:455-68.

23. Galanter J, Choudhry S, Eng C, Nazario S, Rodríguez-Santana JR, Casal J, Torres-Palacios A, Salas J, Chapela R, Watson HG, Meade K, LeNoir M, Rodríguez-Cintrón W, Avila PC, Burchard EG. ORMDL3 gene is associated with asthma in three ethnically diverse populations. Am J Respir Crit Care Med. 2008;177:1194-200.

24. Tavendale R, Macgregor DF, Mukhopadhyay S, Palmer CNA. A polymorphism controlling ORMDL3 expression is associated with asthma that is poorly controlled by current medications. J Allergy Clin Immunol. 2008;121:860-3.

25. Madore A-M, Tremblay K, Hudson TJ, Laprise C. Replication of an association between 17q21 SNPs and asthma in a French-Canadian familial collection. Human genetics. 2008;123:93-5.

26. Hancock DB, Romieu I, Shi M, Sienra-Monge J-J, Wu H, Chiu GY, Li H, del Rio-Navarro BE, Willis-Owen SAG, Willis-Owens SAG, Weiss ST, Raby BA, Gao H, Eng C, Chapela R, Burchard EG, Tang H, Sullivan PF, London SJ. Genome-wide association study implicates chromosome 9q21.31 as a susceptibility locus for asthma in mexican children. PLoS Genet. 2009;5:e1000623.

27. Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, Wilk JB, Willis-Owen SAG, Klanderman B, Lasky-Su J, Lazarus R, Murphy AJ, Soto-Quiros ME, Avila L, Beaty T, Mathias RA, Ruczinski I, Barnes KC, Celedón JC, Cookson WOC, Gauderman WJ, Gilliland FD, Hakonarson H, Lange C, Moffatt MF, O'Connor GT, Raby BA, Silverman EK, Weiss ST. Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. Am J Hum Genet. 2009;84:581-93.

466 28. Bisgaard H, Bønnelykke K, Sleiman PM, Brasholt M, Chawes B, Kreiner-Møller E, Stage M,
467 Kim C, Tavendale R, Baty F. Chromosome 17q21 gene variants are associated with asthma and
468 exacerbations but not atopy in early childhood. Am J Respir Crit Care Med. 2009;179:179-85.
469 29. Wu H, Romieu I, Sienra-Monge J-J, Li H, del Rio-Navarro BE, London SJ. Genetic variation
470 in ORM1-like 3 (ORMDL3) and gasdermin-like (GSDML) and childhood asthma. Allergy.
471 2009;64:629-35.
472 30. Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, Meyers DA, Bleecker ER. Genome-
473 wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. J Allergy
474 Clin Immunol. 2010;125:328-335.e11.
475 31. Sleiman PMA, Flory J, Imielinski M, Bradfield JP, Annaiah K, Willis-Owen SAG, Wang K,
476 Rafaels NM, Michel S, Bonnelykke K, Zhang H, Kim CE, Frackelton EC, Glessner JT, Hou C,
477 Otieno FG, Santa E, Thomas K, Smith RM, Glaberson WR, Garris M, Chiavacci RM, Beaty TH,
478 Ruczinski I, Orange JS, Orange JM, Allen J, Spergel JM, Grundmeier R, Mathias RA, Christie
479 JD, von Mutius E, Cookson WOC, Kabesch M, Moffatt MF, Grunstein MM, Barnes KC, Devoto
480 M, Magnusson M, Li H, Grant SFA, Bisgaard H, Hakonarson H. Variants of DENND1B
481 associated with asthma in children. N Engl J Med. 2010;362:36-44.
482 32. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, von Mutius E, Farrall
483 M, Lathrop M, Cookson WOCM. A large-scale, consortium-based genomewide association
484 study of asthma. N Engl J Med. 2010;363:1211-21.
485 33. Ferreira MA, Matheson MC, Duffy DL, Marks GB, Hui J, Le Souëf P, Danoy P, Baltic S,
486 Nyholt DR, Jenkins M. Identification of IL6R and chromosome 11q13. 5 as risk loci for asthma.
487 The Lancet. 2011;378:1006-14.
488 34. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, Himes
489 BE, Levin AM, Mathias RA, Hancock DB, Baurley JW, Eng C, Stern DA, Celedón JC, Rafaels
490 N, Capurso D, Conti DV, Roth LA, Soto-Quiros M, Togias A, Li X, Myers RA, Romieu I, Van
491 Den Berg DJ, Hu D, Hansel NN, Hernandez RD, Israel E, Salam MT, Galanter J, Avila PC,
492 Avila L, Rodriquez-Santana JR, Chapela R, Rodriguez-Cintron W, Diette GB, Adkinson NF,
493 Abel RA, Ross KD, Shi M, Faruque MU, Dunston GM, Watson HR, Mantese VJ, Ezurum SC,
494 Liang L, Ruczinski I, Ford JG, Huntsman S, Chung KF, Vora H, Li X, Calhoun WJ, Castro M,
495 Sienra-Monge JJ, del Rio-Navarro B, Deichmann KA, Heinzmann A, Wenzel SE, Busse WW,
496 Gern JE, Lemanske RF, Beaty TH, Bleecker ER, Raby BA, Meyers DA, London SJ, Gilliland

497 FD, Burchard EG, Martinez FD, Weiss ST, Williams LK, Barnes KC, Ober C, Nicolae DL.
498 Meta-analysis of genome-wide association studies of asthma in ethnically diverse North
499 American populations. Nat Genet. 2011;43:887-92.
500 35. Swinburne JE, Bogle H, Klukowska-Rötzler J, Drögemüller M, Leeb T, Temperton E, Dolf
501 G, Gerber V. A whole-genome scan for recurrent airway obstruction in Warmblood sport horses
502 indicates two positional candidate regions. Mamm Genome. 2009;20:504-15.
503 36. Klukowska-Rötzler J, Swinburne JE, Drögemüller C, Dolf G, Janda J, Leeb T, Gerber V. The
504 interleukin 4 receptor gene and its role in recurrent airway obstruction in Swiss Warmblood
505 horses. Anim Genet. 2012;43:450-3.
506 37. Jost U, Klukowska-Rötzler J, Dolf G, Swinburne JE, Ramseyer A, Bugno M, Burger D, Blott
507 S, Gerber V. A region on equine chromosome 13 is linked to recurrent airway obstruction in
508 horses. Equine Vet J. 2007;39:236-41.
509 38. Klukowska-Rötzler J, Gerber V, Leeb T. Association analysis of SNPs in the IL21R gene
510 with recurrent airway obstruction (RAO) in Swiss Warmblood horses. Anim Genet.
511 2012;43:475-6.
512 39. Tessier L, Côté O, Clark ME, Viel L, Diaz-Méndez A, Anders S, Bienzle D. Impaired
513 response of the bronchial epithelium to inflammation characterizes severe equine asthma. BMC
514 Genomics. 2017;18:708.
515 40. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear TL, Adelson DL,
516 Bailey E, Bellone RR. Genome sequence, comparative analysis, and population genetics of the
517 domestic horse. Science. 2009;326:865-7.
518 41. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
519 Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15-21.
520 42. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,
521 Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce
522 framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297-
523 303.
524 43. Guo Y, Ding X, Shen Y, Lyon GJ, Wang K. SeqMule: automated pipeline for analysis of
525 human exome/genome sequencing data. Sci Rep. 2015;5:srep14283.

526 44. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the
527 consequences of genomic variants with the Ensembl API and SNP Effect Predictor.
528 Bioinformatics. 2010;26:2069-70.
529 45. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res.
530 2001;11:863-74.
531 46. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein
532 function. Genome Res. 2002;12:436-46.
533 47. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic
534 Acids Res. 2003;31:3812-4.
535 48. Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function.
536 Annu Rev Genomics Hum Genet. 2006;7:61-80.
537 49. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on
538 protein function using the SIFT algorithm. Nat Protoc. 2009;4:1073-81.
539 50. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense
540 mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit7.20.
541 51. Hecht M, Bromberg Y, Rost B. News from the protein mutability landscape. J Mol Biol.
542 2013;425:3937-48.
543 52. Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function.
544 Nucleic Acids Res. 2007;35:3823-35.
545 53. Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence variants.
546 BMC Genomics. 2015;16 Suppl 8:S1.
547 54. Quinn EM, Cormican P, Kenny EM, Hill M, Anney R, Gill M, Corvin AP, Morris DW.
548 Development of strategies for SNP detection in RNA-seq data: application to lymphoblastoid
549 cell lines and evaluation using 1000 Genomes data. PLoS One. 2013;8:e58815.
550 55. Piskol R, Ramaswami G, Li JB. Reliable identification of genomic variants from RNA-seq
551 data. Am J Hum Genet. 2013;93:641-51.
552 56. Sheng Q, Zhao S, Li C-I, Shyr Y, Guo Y. Practicability of detecting somatic point mutation
553 from RNA high throughput sequencing data. Genomics. 2016;107:163-9.
554 57. Guo Y, Zhao S, Sheng Q, Samuels DC, Shyr Y. The discrepancy among single nucleotide
555 variants detected by DNA and RNA high throughput sequencing data. BMC Genomics.
556 2017;18:690.

557 58. Li X, Howard TD, Moore WC, Ampleford EJ, Li H, Busse WW, Calhoun WJ, Castro M,
558 Chung KF, Erzurum SC, Fitzpatrick AM, Gaston B, Israel E, Jarjour NN, Teague WG, Wenzel
559 SE, Peters SP, Hawkins GA, Bleecker ER, Meyers DA. Importance of hedgehog interacting
560 protein and other lung function genes in asthma. J Allergy Clin Immunol. 2011;127:1457-65.
561 59. Flanagan SE, Patch A-M, Ellard S. Using SIFT and PolyPhen to predict loss-of-function and
562 gain-of-function mutations. Genet Test Mol Biomarkers. 2010;14:533-7.
563 60. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X. Comparison and integration
564 of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing
565 studies. Hum Mol Genet. 2015;24:2125-37.
566 61. Thusberg J, Vihinen M. Pathogenic or not? And if so, then how? Studying the effects of
567 missense mutations using bioinformatics methods. Hum Mutat. 2009;30:703-14.
568 62. Thumberger T, Hagenlocher C, Tisler M, Beyer T, Tietze N, Schweickert A, Feistel K, Blum
569 M. Ciliary and non-ciliary expression and function of PACRG during vertebrate development.
570 Cilia. 2012;1:13.
571 63. West AB, Lockhart PJ, O'Farell C, Farrer MJ. Identification of a novel gene linked to parkin
572 via a bi-directional promoter. J Mol Biol. 2003;326:11-9.
573 64. Mizuno K, Dymek EE, Smith EF. Microtubule binding protein PACRG plays a role in
574 regulating specific ciliary dyneins during microtubule sliding. Cytoskeleton. 2016;73:703-11.
575 65. Wilson GR, Wang HX, Egan GF, Robinson PJ, Delatycki MB, O'Bryan MK, Lockhart PJ.
576 Deletion of the Parkin co-regulated gene causes defects in ependymal ciliary motility and
577 hydrocephalus in the quakingviable mutant mouse. Hum Mol Genet. 2010;19:1593-602.
578 66. Ikeda T. Parkin-co-regulated gene (PACRG) product interacts with tubulin and microtubules.
579 FEBS Lett. 2008;582:1413-8.
580 67. Scholtens S, Postma DS, Moffatt MF, Panasevich S, Granell R, Henderson AJ, Melén E,
581 Nyberg F, Pershagen G, Jarvis D, Ramasamy A, Wjst M, Svanes C, Bouzigon E, Demenais F,
582 Kauffmann F, Siroux V, von Mutius E, Ege MJ, Braun-Fahrländer C, Genuneit J, group GS,
583 Brunekreef B, Smit HA, Wijga AH, Kerkhof M, Curjuric I, Imboden M, Thun GA, Probst-
584 Hensch N, Freidin MB, Bragina EI, Deev IA, Puzyrev VP, Daley D, Park J, Becker A, Chan-
585 Yeung M, Kozyrskyj AL, Pare P, Marenholz I, Lau S, Keil T, Lee Y-A, Kabesch M, Wijmenga
586 C, Franke L, Nolte IM, Vonk J, Kumar A, Farrall M, Cookson WO, Strachan DP, Koppelman

587 GH, Boezen HM. Novel childhood asthma genes interact with in utero and early-life tobacco
588 smoke exposure(). J Allergy Clin Immunol. 2014;133:885-8.
589 68. Betts MJ, Russell RB. Amino Acid Properties and Consequences of Substitutions. In:
590 Bioinformatics for Geneticists. John Wiley & Sons, Ltd; 2003. p. 289-316.
591 69. Faisst AM, Alvarez-Bolado G, Treichel D, Gruss P. Rotatin is a novel gene required for axial
592 rotation and left-right specification in mouse embryos. Mech Dev. 2002;113:15-28.
593 70. Chen H-Y, Wu C-T, Tang C-JC, Lin Y-N, Wang W-J, Tang TK. Human microcephaly
594 protein RTTN interacts with STIL and is required to build full-length centrioles. Nat Commun.
595 2017;8:247.
596 71. Kheradmand Kia S, Verbeek E, Engelen E, Schot R, Poot RA, de Coo IFM, Lequin MH,
597 Poulton CJ, Pourfarzad F, Grosveld FG, Brehm A, de Wit MCY, Oegema R, Dobyns WB,
598 Verheijen FW, Mancini GMS. RTTN mutations link primary cilia function to organization of the
599 human cerebral cortex. Am J Hum Genet. 2012;91:533-40.
600 72. Udina IG, Kordicheva SI, Pospelov LE, Malenko AF, Gergert VI, Pospelov AL, Matrashkin
601 AG, Kyzyl-Ool MM, Nachin AA, Zhivotovskiĭ LA. [Study of the polymorphic markers--the
602 PARK2 and PACRG genes due to the incidence of pulmonary tuberculosis in two districts of the
603 Republic of Tyva]. Probl Tuberk Bolezn Legk. 2007:27-9.
604 73. Bragina EY, Tiys ES, Rudko AA, Ivanisenko VA, Freidin MB. Novel tuberculosis
605 susceptibility candidate genes revealed by the reconstruction and analysis of associative
606 networks. Infect Genet Evol. 2016;46:118-23.
607 74. Majewski J, Ott J. Amino acid substitutions in the human genome: evolutionary implications
608 of single nucleotide polymorphisms. Gene. 2003;305:167-73.
609 75. Betts MJ, Russell RB. Amino acid properties and consequences of substitutions.
610 Bioinformatics for geneticists. 2003;317:289.
611 76. Khan S, Vihinen M. Spectrum of disease-causing mutations in protein secondary structures.
612 BMC Struct Biol. 2007;7:56.
613 77. Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. Nat Rev
614 Mol Cell Biol. 2017;18:533-47.
615 78. Szymanska K, Johnson CA. The transition zone: an essential functional compartment of cilia.
616 Cilia. 2012;1:10.

79. Eshel D, Priel Z. Characterization of metachronal wave of beating cilia on frog's palate epithelium in tissue culture. J Physiol. 1987;388:1-8.
80. Gheber L, Priel Z. On metachronism in ciliary systems: a model describing the dependence of the metachronal wave properties on the intrinsic ciliary parameters. Cell Motil Cytoskeleton. 1990;16:167-81.
81. Steffen W, Linck RW. Evidence for tektins in centrioles and axonemal microtubules. Proc Natl Acad Sci U S A. 1988;85:2643-7.
82. Linck RW, Norrander JM. Protofilament ribbon compartments of ciliary and flagellar microtubules. Protist. 2003;154:299.
83. Kaup FJ, Drommer W, Deegen E. Ultrastructural findings in horses with chronic obstructive pulmonary disease (COPD). I: Alterations of the larger conducting airways. Equine Vet J. 1990;22:343-8.
84. Huangfu D, Liu A, Rakeman AS, Murcia NS, Niswander L, Anderson KV. Hedgehog signalling in the mouse requires intraflagellar transport proteins. Nature. 2003;426:83-7.
85. Goetz SC, Anderson KV. The primary cilium: a signalling centre during vertebrate development. Nat Rev Genet. 2010;11:331-44.
86. Gavino C, Richard S. Patched1 haploinsufficiency impairs ependymal cilia function of the quaking viable mice, leading to fatal hydrocephalus. Mol Cell Neurosci. 2011;47:100-7.
87. Ikeda K, Ikeda T, Morikawa K, Kamiya R. Axonemal localization of Chlamydomonas PACRG, a homologue of the human Parkin-coregulated gene product. Cell Motil Cytoskeleton. 2007;64:814-21.
88. Lechtreck K-F, Luro S, Awata J, Witman GB. HA-tagging of putative flagellar proteins in Chlamydomonas reinhardtii identifies a novel protein of intraflagellar transport complex B. Cell Motil Cytoskeleton. 2009;66:469-82.
89. Dawe HR, Farr H, Portman N, Shaw MK, Gull K. The Parkin co-regulated gene product, PACRG, is an evolutionarily conserved axonemal protein that functions in outer-doublet microtubule morphogenesis. J Cell Sci. 2005;118:5421-30.
90. Li Y, Xiao X, Ji X, Liu B, Amos CI. RNA-seq analysis of lung adenocarcinomas reveals different gene expression profiles between smoking and nonsmoking patients. Tumour Biol. 2015;36:8993-9003.

647 91. Imai Y, Soda M, Murakami T, Shoji M, Abe K, Takahashi R. A product of the human gene
648 adjacent to parkin is a component of Lewy bodies and suppresses Pael receptor-induced cell
649 death. J Biol Chem. 2003;278:51901-10.
650 92. Valley CC, Cembran A, Perlmutter JD, Lewis AK, Labello NP, Gao J, Sachs JN. The
651 methionine-aromatic motif plays a unique role in stabilizing protein structure. Journal of
652 Biological Chemistry. 2012;287:34979-91.
653 93. Wolfenden R. Experimental measures of amino acid hydrophobicity and the topology of
654 transmembrane and globular proteins. J Gen Physiol. 2007;129:357-62.
655 94. Brosnan JT, Brosnan ME. The sulfur-containing amino acids: an overview. J Nutr.
656 2006;136:1636S-40S.
657 95. Kazemi-Esfarjani P, Beitel LK, Trifiro M, Kaufman M, Rennie P, Sheppard P, Matusik R,
658 Pinsky L. Substitution of valine-865 by methionine or leucine in the human androgen receptor
659 causes complete or partial androgen insensitivity, respectively with distinct androgen receptor
660 phenotypes. Molecular Endocrinology. 1993;7:37-46.
661

LEGENDS

Figure 1. Substitution variants in *PACRG* (A) and *RTTN* (B) genes. Diagrams show position of gene (thick green line), mRNA (red line) and coding region (yellow line). Whole gene (upper) and close-up view surrounding the substitution variant (lower) are included with gene sequence at the bottom. Location of the substitution variant is indicated in blue for NCBI and Ensembl databases.

Figure 2. Alignment of wild type (WT) and altered (alt) PACRG proteins with associated predicted hydrophobicity and isoelectric point. Replacement of methionine for valine at position 182 changes a beta strand to an alpha helix at position 187 and 188. Alpha helices (pink), coils (gray line), turns (blue arrows) and beta strands (yellow arrows).

Figure 3. Alignment of WT and altered RTTN protein with predicted hydrophobicity and isoelectric point. Replacement of arginine with tryptophan at position 1812 changes alpha helix to beta strand at position 1816. Alpha helices (pink), coils (gray line), turns (blue arrows) and beta strands (yellow arrows).

Figure 4. Expression of PACRG (A, B) and RTTN (C, D) in asthmatic and non-asthmatic horses in counts-per-million (CPM; y-axis) pre- and post-challenge. PACRG expression varied from ~30.7 to 66.3 CPM in asthmatic horses (A) and ~25.1 to 65.6 CPM on non-asthmatic horses (B), while RTTN expression varied from ~11.5 to 24.9 CPM and ~13.1 to 29.9 CPM in asthmatic (C) and non-asthmatic (D) horses, respectively.

Figure 5. Comparison of GATK substitution variant calls and Sanger sequencing results for *PACRG* in asthmatic (A) and non-asthmatic (B) horses. For both groups, the bar graph indicates the IGV count for each allele (A-red, G-yellow), horse and condition. Below the bar graph is the GATK variant call, the electropherogram of the Sanger sequence, and agreement. (A) Four asthmatic horses (1, 2, 5, 6) had heterozygous alleles [A/G] and two (3 and 4) were homozygous for the altered allele [A/A]. (B) In non-asthmatic horses, one horse (6) had heterozygous alleles [A/G], three horses (1, 2 and 4) were homozygous for the wild-type allele [G/G] and three horses

(3, 5 and 7) were homozygous for the altered allele [A/A]. All genotypes were consistent across horses and methods. DNA was not available for non-asthmatic horse 1.

Figure 6. Comparison of GATK substitution variant calls and Sanger sequencing results for *RTTN* in asthmatic (A) and non-asthmatic (B) horses. Details as in Fig. 5. (A) Four asthmatic horses (3-6) had heterozygous alleles [A/T] and two (1 and 2) were homozygous for the altered allele [A/A]. Genotypes were consistent across horses and methods. (B) In non-asthmatic horses, one (5) had heterozygous [A/T] alleles, four horses had homozygous wild type [T/T] alleles, and two horses (1 and 4) were inconsistently identified as homozygous wild type and heterozygous in different samples. Sanger sequencing confirmed the genotype of horse 4 as heterozygous. DNA was not available for non-asthmatic horse 1.

Figure 7. Alignment of PACRG (A) and RTTN (B) Sanger sequences for 10 asthmatic and 14 non-asthmatic horses with the reference genome. (A) For PACRG, 8 asthmatic horses (80%) were heterozygous [A/G], 2 (20%) were homozygous-altered [A/A] and none was homozygous wild-type [G/G]. Six non-asthmatic horses (43%) were heterozygous [A/G], 5 (36%) were homozygous-altered [A/A] and 3 (21%) were homozygous wild type [G/G]. (B) For RTTN, in the asthmatic group, there were 5 (50%) heterozygous [A/T], 2 (20%) homozygous-altered [A/A] and 3 (30%) homozygous wild type [T/T] genotypes. In the non-asthmatic group, 4 (29%) were heterozygous [A/G], 10 (71%) were homozygous wild type [T/T], and none had the homozygous-altered [A/A] genotype.

Figure 1

Substitution variants in *PACRG* (A) and *RTTN* (B) genes.

Diagrams show position of gene (thick green line), mRNA (red line) and coding region (yellow line). Whole gene (upper) and close-up view surrounding the substitution variant (lower) are included with gene sequence at the bottom. Location of the substitution variant is indicated in blue for NCBI and Ensembl databases.

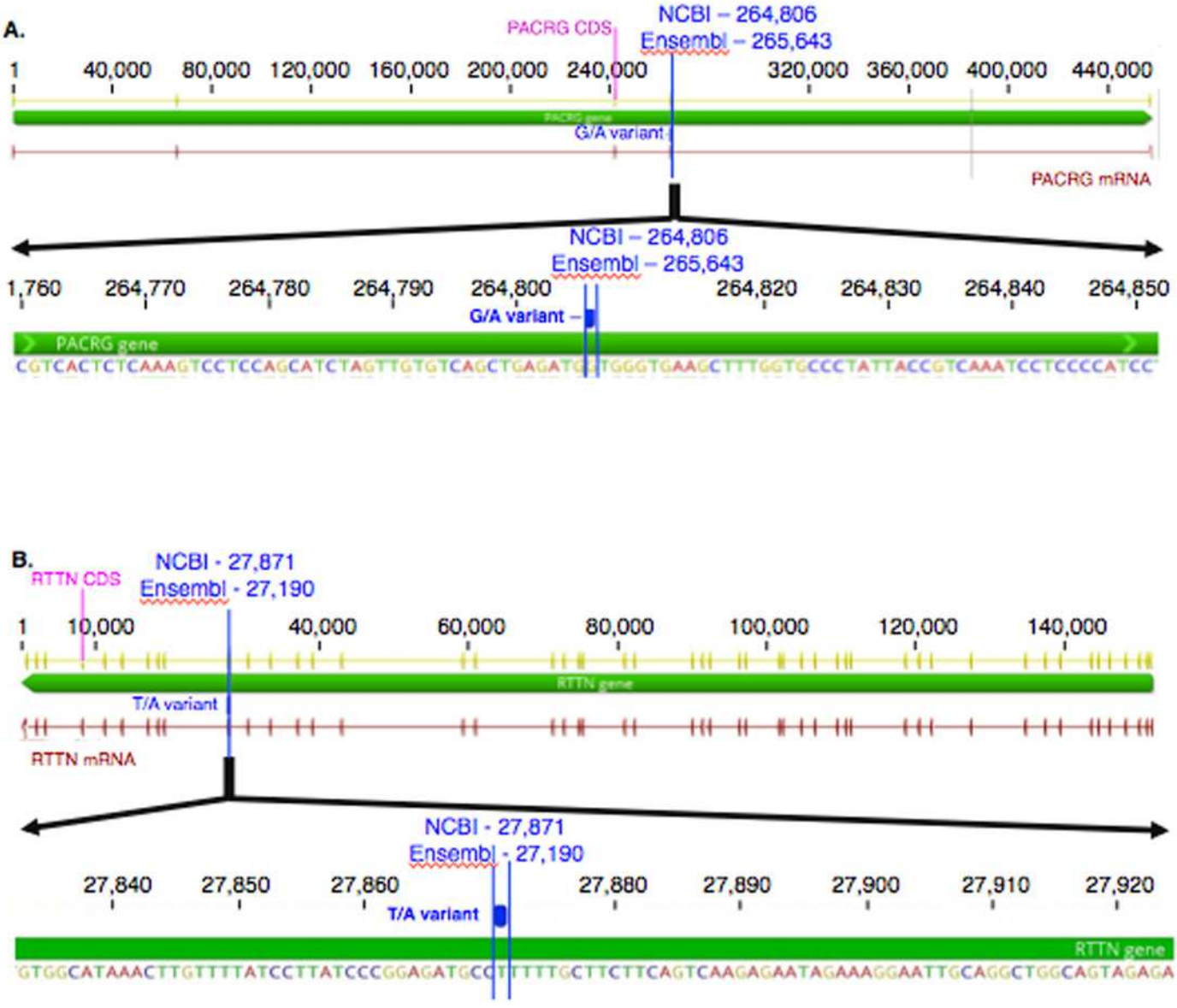


Figure 2

Alignment of wild type (WT) and mutant (mut) PACRG proteins with associated predicted hydrophobicity and isoelectric point.

Replacement of methionine for valine at position 182 changes a beta strand to an alpha helix at position 187 and 188. Alpha helices (pink), coils (gray line), turns (blue arrows) and beta strands (yellow arrows).

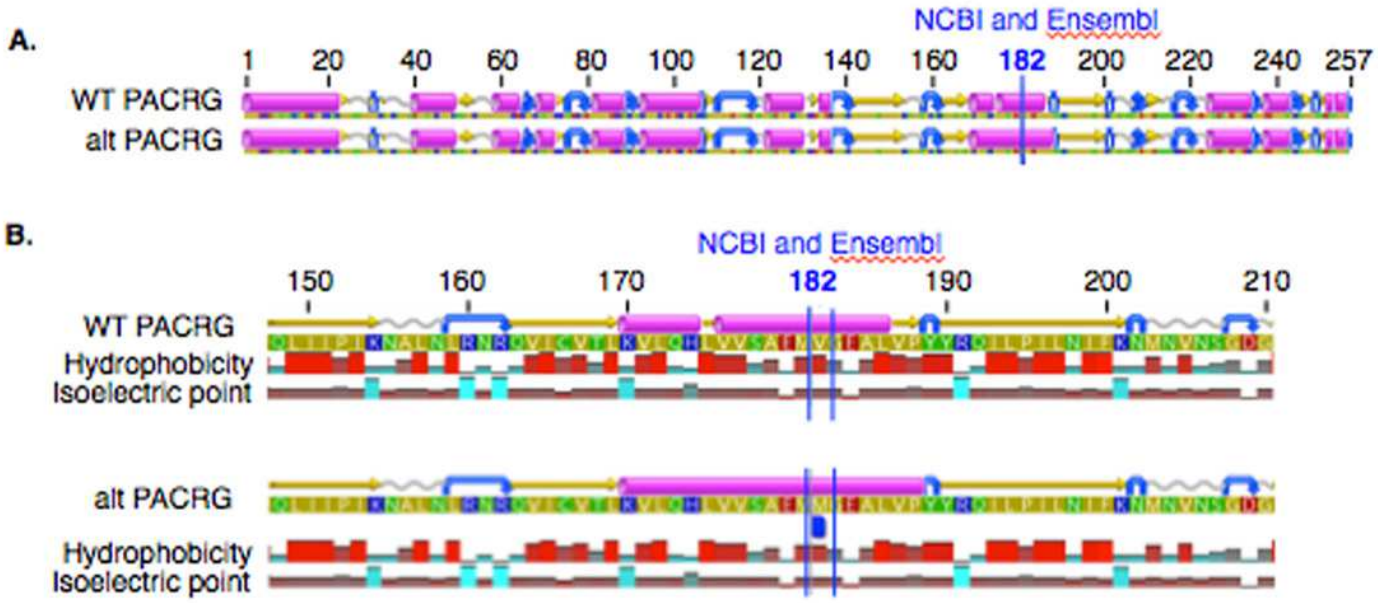


Figure 3

Alignment of WT and mutant RTTN protein with predicted hydrophobicity and isoelectric point.

Replacement of arginine with tryptophan at position 1812 changes alpha helix to beta strand at position 1816. Alpha helices (pink), coils (gray line), turns (blue arrows) and beta strands (yellow arrows).

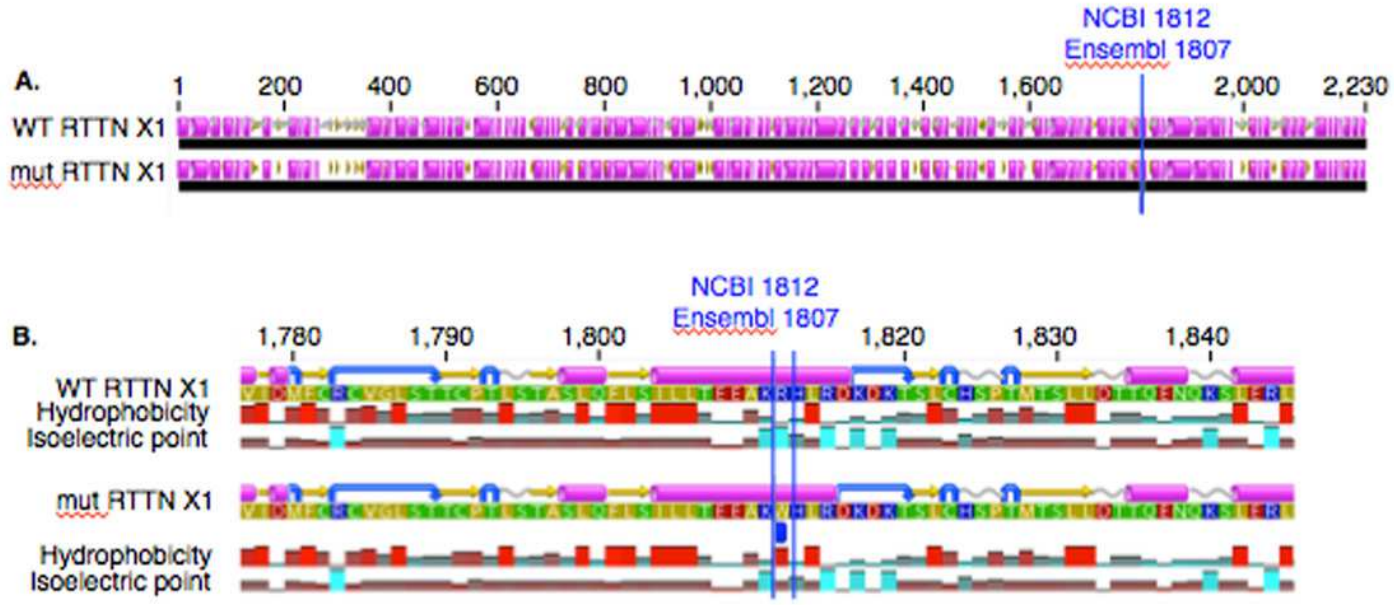


Figure 4

Expression of PACRG (A, B) and RTTN (C, D) in asthmatic and non-asthmatic horses in counts-per-million (CPM; y-axis) pre- and post-challenge.

PACRG expression varied from ~30.7 to 66.3 CPM in asthmatic horses (A) and ~25.1 to 65.6 CPM on non-asthmatic horses (B), while RTTN expression varied from ~11.5 to 24.9 CPM and ~13.1 to 29.9 CPM in asthmatic (C) and non-asthmatic (D) horses, respectively.

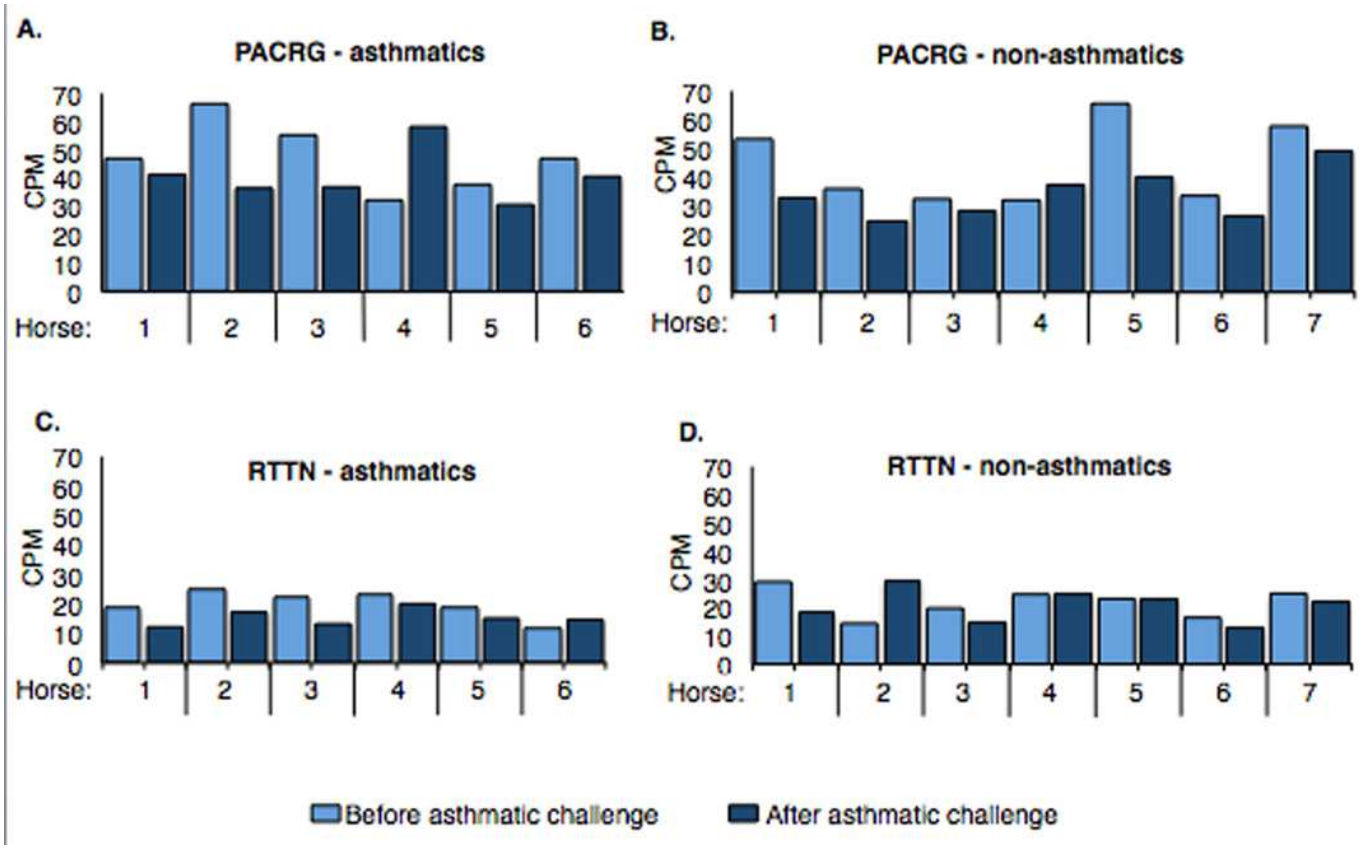


Figure 5

Comparison of GATK variant calls and Sanger sequencing results for *PACRG* in asthmatic (A) and non-asthmatic (B) horses.

For both groups, the bar graph indicates the IGV count for each allele (A-red, G-yellow), horse and condition. Below the bar graph is the GATK variant call, the electropherogram of the Sanger sequence, and agreement. (A) Four asthmatic horses (1, 2, 5, 6) had heterozygous alleles [A/G] and two (3 and 4) were homozygous for the mutant allele [A/A]. (B) In non-asthmatic horses, one horse (6) had heterozygous alleles [A/G], three horses (1, 2 and 4) were homozygous for the wild-type allele [G/G] and three horses (3, 5 and 7) were homozygous for the mutant allele [A/A]. All genotypes were consistent across horses and methods. DNA was not available for non-asthmatic horse 1.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3491v2 | CC BY 4.0 Open Access | rec: 29 Aug 2018, publ: 29 Aug 2018

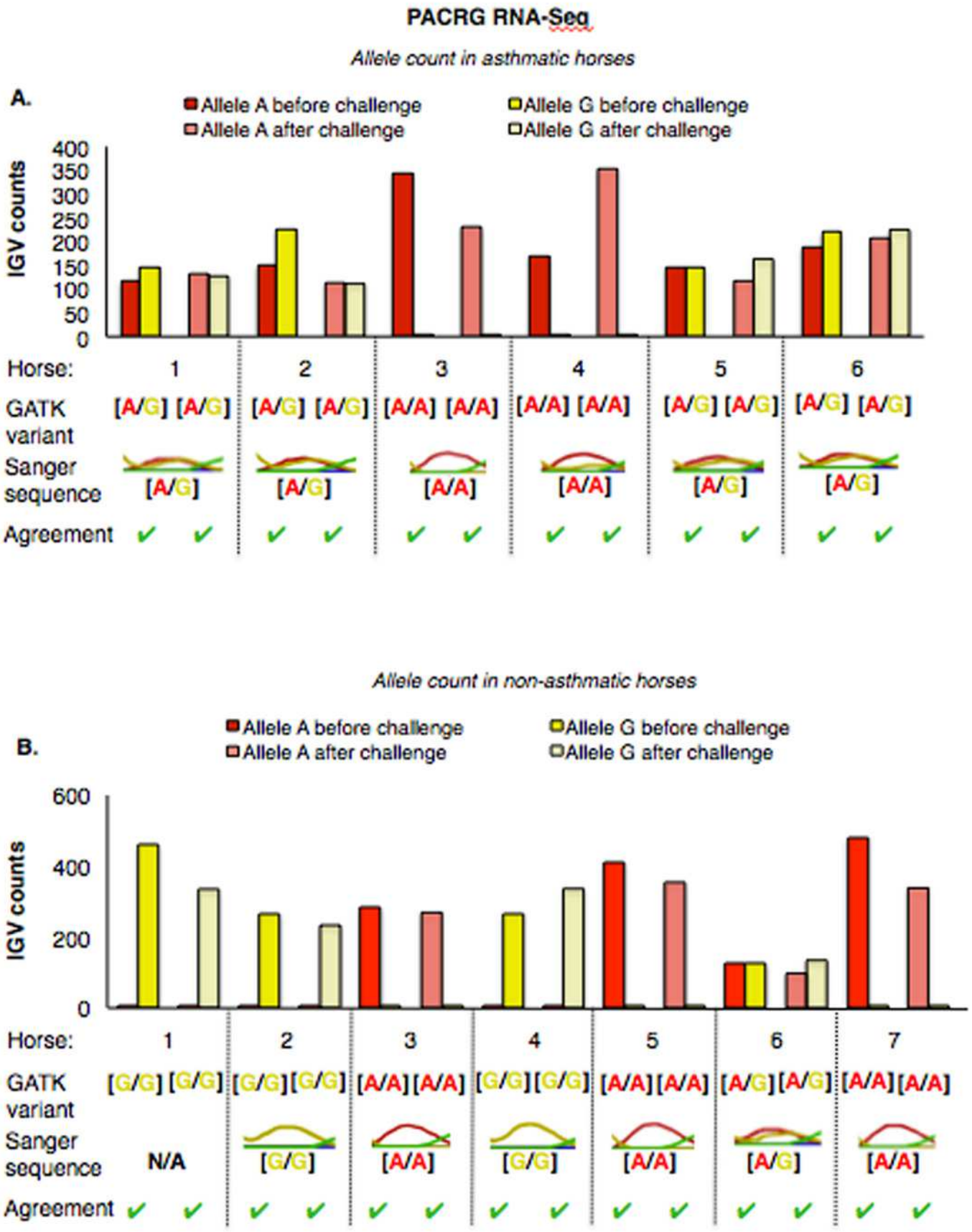


Figure 6

Comparison of GATK variant calls and Sanger sequencing results for *RTTN* in asthmatic (A) and non-asthmatic (B) horses.

Details as in Fig. 6. (A) Four asthmatic horses (3-6) had heterozygous alleles [A/T] and two (1 and 2) were homozygous for the mutant allele [A/A]. Genotypes were consistent across horses and methods. (B) In non-asthmatic horses, one (5) had heterozygous [A/T] alleles, four horses had homozygous wild type [T/T] alleles, and two horses (1 and 4) were inconsistently identified as homozygous wild type and heterozygous in different samples. Sanger sequencing confirmed the genotype of horse 4 as heterozygous. DNA was not available for non-asthmatic horse 1.

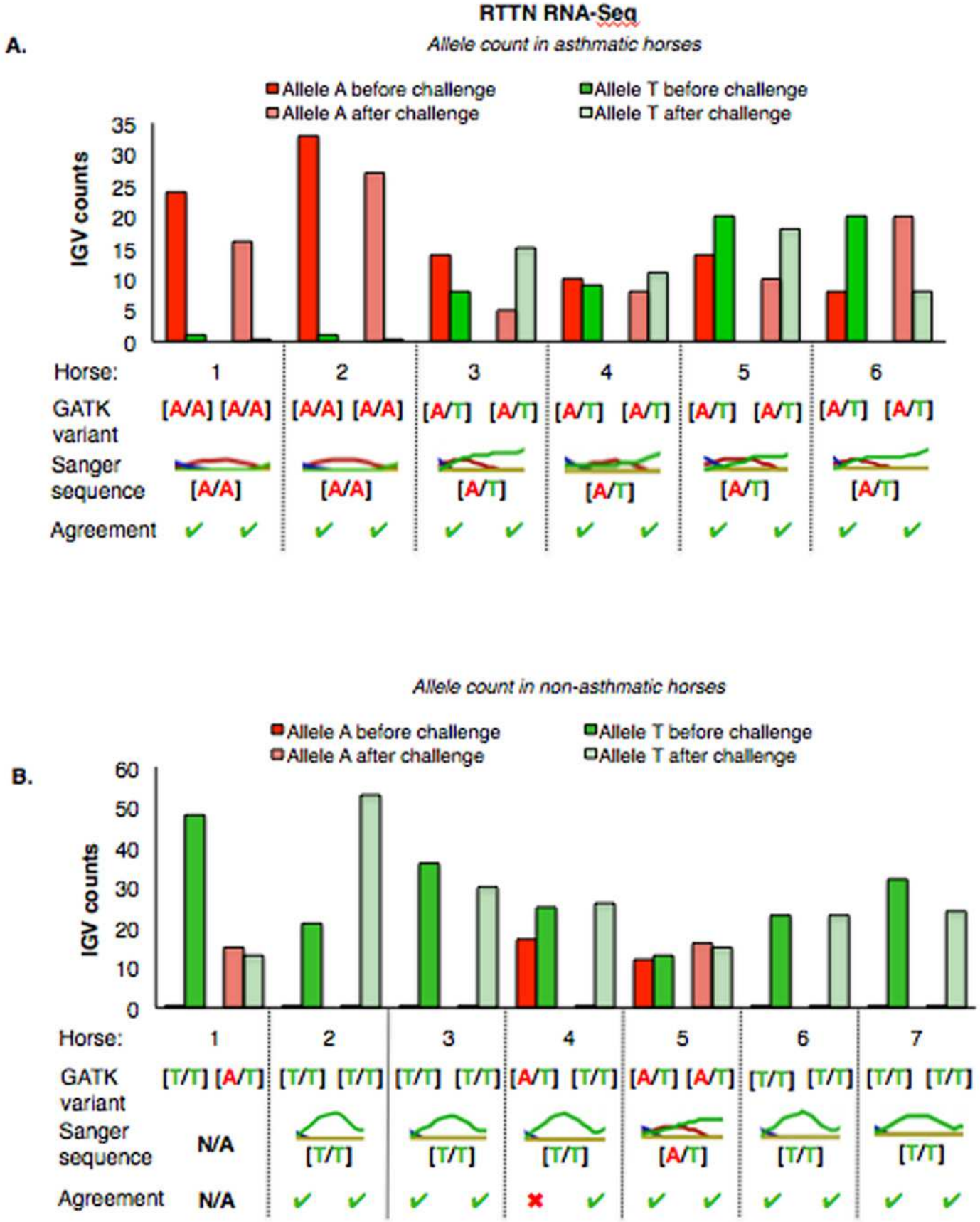
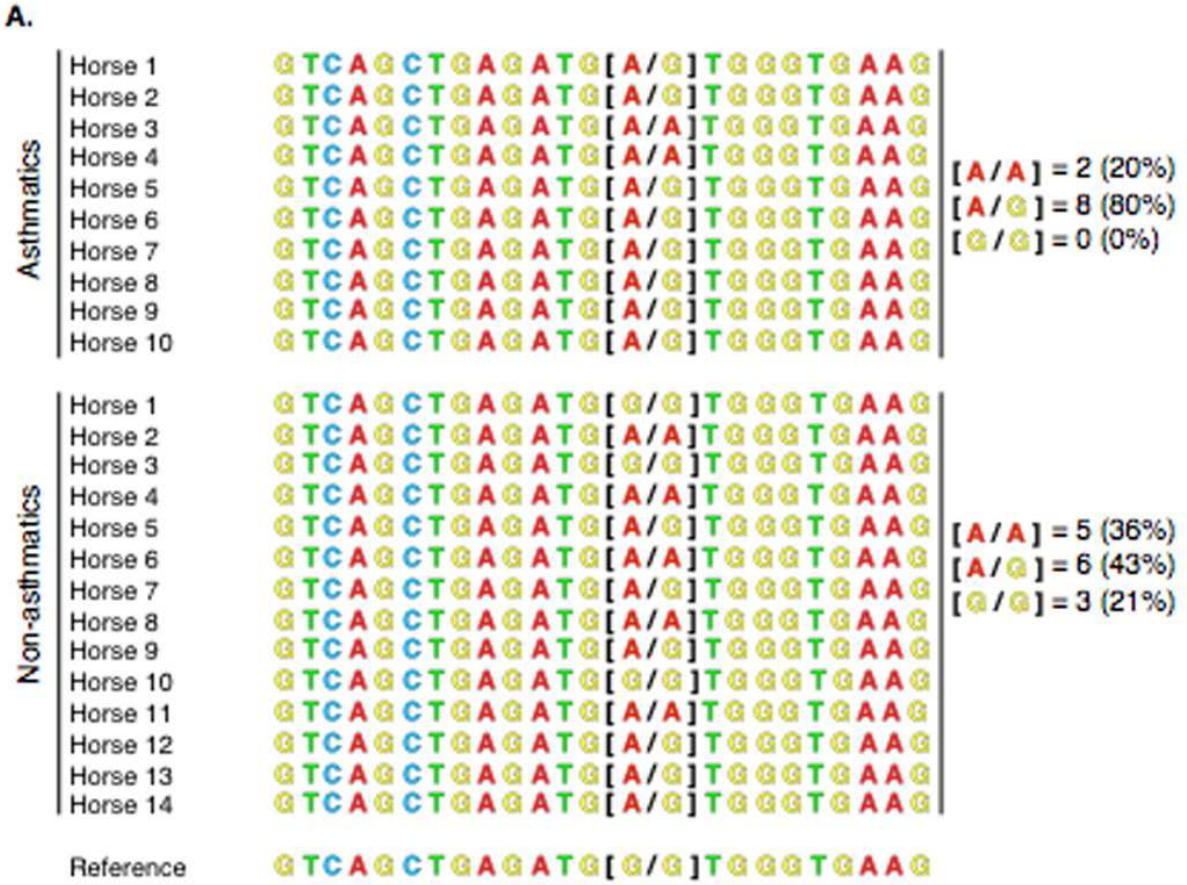
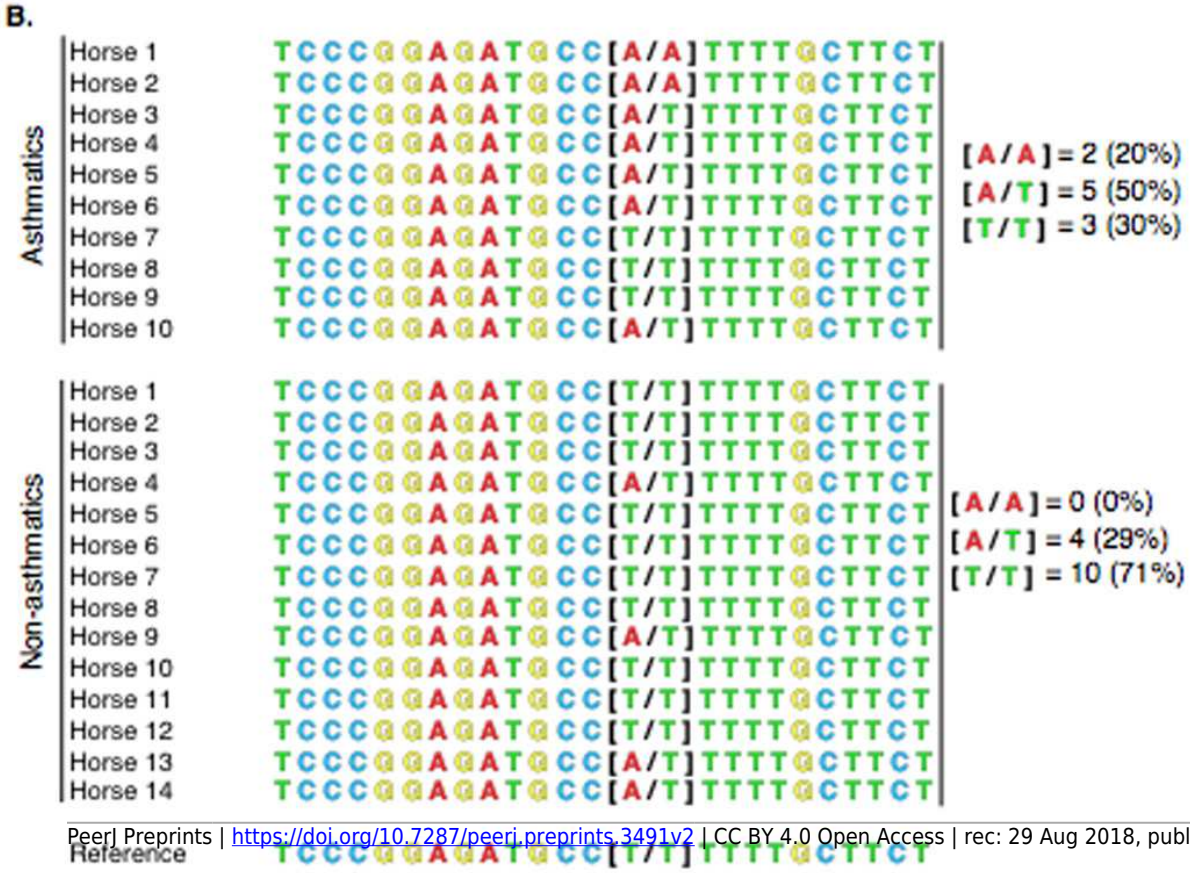

A.
RTTN RNA-Seq
Allele count in asthmatic horses
Allele A before challenge
Allele A after challenge
Allele T before challenge
Allele T after challenge
IGV counts
35
30
25
20
15
10
5
0
Horse: 1 2 3 4 5 6
GATK variant
[A/A] [A/A] [A/A] [A/A] [A/T] [A/T] [A/T] [A/T] [A/T] [A/T] [A/T] [A/T]
Sanger sequence
[A/A] [A/A] [A/T] [A/T] [A/T] [A/T]
Agreement
✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔
B.
Allele count in non-asthmatic horses
Allele A before challenge
Allele A after challenge
Allele T before challenge
Allele T after challenge
IGV counts
60
50
40
30
20
10
0
Horse: 1 2 3 4 5 6 7
GATK variant
[T/T] [A/T] [T/T] [T/T] [T/T] [T/T] [A/T] [T/T] [A/T] [A/T] [T/T] [T/T] [T/T] [T/T]
Sanger sequence
N/A [T/T] [T/T] [T/T] [A/T] [T/T] [T/T]
Agreement
N/A ✔ ✔ ✔ ✔ ✖ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔

Figure 7

Alignment of PACRG (A) and RTTN (B) Sanger sequences for 10 asthmatic and 14 non-asthmatic horses with the reference genome.

(A) For PACRG, 8 asthmatic horses (80%) were heterozygous [A/G], 2 (20%) were homozygous mutant [A/A] and none was homozygous wild-type [G/G]. Six non-asthmatic horses (43%) were heterozygous [A/G], 5 (36%) were homozygous mutant [A/A] and 3 (21%) were homozygous wild type [G/G]. (B) For RTTN, in the asthmatic group, there were 5 (50%) heterozygous [A/T], 2 (20%) homozygous mutant [A/A] and 3 (30%) homozygous wild type [T/T] genotypes. In the non-asthmatic group, 4 (29%) were heterozygous [A/G], 10 (71%) were homozygous wild type [T/T], and none had the homozygous mutant [A/A] genotype.


A.
Asthmatics
Horse 1
Horse 2
Horse 3
Horse 4
Horse 5
Horse 6
Horse 7
Horse 8
Horse 9
Horse 10
GTCAGCTGAGATG[A/G]TGGGTGAAG
GTCAGCTGAGATG[A/G]TGGGTGAAG
GTCAGCTGAGATG[A/A]TGGGTGAAG
GTCAGCTGAGATG[A/A]TGGGTGAAG
GTCAGCTGAGATG[A/G]TGGGTGAAG
GTCAGCTGAGATG[A/G]TGGGTGAAG
GTCAGCTGAGATG[A/G]TGGGTGAAG
GTCAGCTGAGATG[A/G]TGGGTGAAG
GTCAGCTGAGATG[A/G]TGGGTGAAG
GTCAGCTGAGATG[A/G]TGGGTGAAG
[A/A] = 2 (20%)
[A/G] = 8 (80%)
[G/G] = 0 (0%)
Non-asthmatics
Horse 1
Horse 2
Horse 3
Horse 4
Horse 5
Horse 6
Horse 7
Horse 8
Horse 9
Horse 10
Horse 11
Horse 12
Horse 13
Horse 14
GTCAGCTGAGATG[G/G]TGGGTGAAG
GTCAGCTGAGATG[A/A]TGGGTGAAG
GTCAGCTGAGATG[G/G]TGGGTGAAG
GTCAGCTGAGATG[A/A]TGGGTGAAG
GTCAGCTGAGATG[A/G]TGGGTGAAG
GTCAGCTGAGATG[A/A]TGGGTGAAG
GTCAGCTGAGATG[A/G]TGGGTGAAG
GTCAGCTGAGATG[A/A]TGGGTGAAG
GTCAGCTGAGATG[A/G]TGGGTGAAG
GTCAGCTGAGATG[G/G]TGGGTGAAG
GTCAGCTGAGATG[A/A]TGGGTGAAG
GTCAGCTGAGATG[A/G]TGGGTGAAG
GTCAGCTGAGATG[A/G]TGGGTGAAG
GTCAGCTGAGATG[A/G]TGGGTGAAG
[A/A] = 5 (36%)
[A/G] = 6 (43%)
[G/G] = 3 (21%)
Reference
GTCAGCTGAGATG[G/G]TGGGTGAAG


B.
Asthmatics
Horse 1
Horse 2
Horse 3
Horse 4
Horse 5
Horse 6
Horse 7
Horse 8
Horse 9
Horse 10
TCCCGGAGATGCC[A/A]TTTTGCTTCT
TCCCGGAGATGCC[A/A]TTTTGCTTCT
TCCCGGAGATGCC[A/T]TTTTGCTTCT
TCCCGGAGATGCC[A/T]TTTTGCTTCT
TCCCGGAGATGCC[A/T]TTTTGCTTCT
TCCCGGAGATGCC[A/T]TTTTGCTTCT
TCCCGGAGATGCC[T/T]TTTTGCTTCT
TCCCGGAGATGCC[T/T]TTTTGCTTCT
TCCCGGAGATGCC[T/T]TTTTGCTTCT
TCCCGGAGATGCC[A/T]TTTTGCTTCT
[A/A] = 2 (20%)
[A/T] = 5 (50%)
[T/T] = 3 (30%)
Non-asthmatics
Horse 1
Horse 2
Horse 3
Horse 4
Horse 5
Horse 6
Horse 7
Horse 8
Horse 9
Horse 10
Horse 11
Horse 12
Horse 13
Horse 14
TCCCGGAGATGCC[T/T]TTTTGCTTCT
TCCCGGAGATGCC[T/T]TTTTGCTTCT
TCCCGGAGATGCC[T/T]TTTTGCTTCT
TCCCGGAGATGCC[A/T]TTTTGCTTCT
TCCCGGAGATGCC[T/T]TTTTGCTTCT
TCCCGGAGATGCC[T/T]TTTTGCTTCT
TCCCGGAGATGCC[T/T]TTTTGCTTCT
TCCCGGAGATGCC[T/T]TTTTGCTTCT
TCCCGGAGATGCC[A/T]TTTTGCTTCT
TCCCGGAGATGCC[T/T]TTTTGCTTCT
TCCCGGAGATGCC[T/T]TTTTGCTTCT
TCCCGGAGATGCC[T/T]TTTTGCTTCT
TCCCGGAGATGCC[A/T]TTTTGCTTCT
TCCCGGAGATGCC[A/T]TTTTGCTTCT
[A/A] = 0 (0%)
[A/T] = 4 (29%)
[T/T] = 10 (71%)
Reference
TCCCGGAGATGCC[T/T]TTTTGCTTCT