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

Variant analysis of RNA sequences in severe equine asthma

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

[1] Pathobiology, University of Guelph

[2] BioAssayWorks, Ijamsville, MD

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. The disease in horses has complex inheritance including both dominant and recessive patterns that are ill defined. This study aimed to determine the utility of RNA-Seq to call gene variants and identify mutations potentially linked to disease. **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 variant calling was performed with Picard tools and Genome Analysis Toolkit (GATK). Coverage was visualized using Integrative Genomic Viewer software and variants were called and filtered using GATK and Ensembl Variant Effect Predictor (VEP) tools. Novel variant selection by VEP was based on 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 selected mutations, the effect of predicted variants on protein function was assessed with Polymorphism Phenotyping (PolyPhen) 2 and Screening for Non-Acceptable Polymorphism (SNAP) 2 softwares. RNA-Seq predicted variants were confirmed in all horses, and investigated in an additional 4 asthmatic and 7 non-asthmatic individuals with PCR and Sanger sequencing. Gene alignment and 3D protein structures were predicted with Geneious software. **Results.** Level of expression across the genome was similar in all individuals. RNA-Seq variant calling and filtering identified with highest confidence mutations in *PACRG* and *RTTN*. Sanger sequencing confirmed that the *PACRG* variant was appropriately identified in all 26 samples while the *RTTN* variant was identified correctly by RNA-Seq in 24 of 26 samples. SIFT and PolyPhen2 indicated both mutations would result in loss of function, and SNAP2 that they would be non-neutral. Amino acid substitutions projected no change of hydrophobicity and isoelectric point in PACRG, a change in both for RTTN; and a slight change in 3D structure for PACRG and RTTN. For *PACRG,* samples from additional individuals confirmed higher frequency of the heterozygous genotype in

asthmatics, while the *RTTN* homozygous mutant phenotype was more prevalent in the asthmatic compared to non-asthmatic group. **Discussion.** RNA-Seq was sensitive and specific for calling gene 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 variants in low coverage samples. The impact of amino acid alterations in PACRG and RTTN proteins are unknown at this point, but their role in structure and function of cilia may warrant further investigation.

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: BioAssay Works LLC, 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. The disease in horses has complex inheritance including both dominant and recessive patterns that are ill defined. This study aimed to determine the utility of RNA-Seq to call gene variants and identify mutations potentially linked to disease.

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 variant calling was performed with Picard tools and Genome Analysis Toolkit (GATK). Coverage was visualized using Integrative Genomic Viewer software and variants were called and filtered using GATK and Ensembl Variant Effect Predictor (VEP) tools. Novel variant selection by VEP was based on 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 selected mutations, the effect of predicted variants on protein function was assessed with Polymorphism Phenotyping (PolyPhen) 2 and Screening for Non-Acceptable Polymorphism (SNAP) 2 softwares. RNA-Seq predicted variants were confirmed in all horses, and investigated in an additional 4 asthmatic and 7 non-asthmatic individuals with PCR and Sanger sequencing. Gene alignment and 3D protein structures were predicted with Geneious software.

Results. Level of expression across the genome was similar in all individuals. RNA-Seq variant calling and filtering identified with highest confidence mutations in *PACRG* and *RTTN*. Sanger sequencing confirmed that the *PACRG* variant was appropriately identified in all 26 samples while the *RTTN* variant was identified correctly by RNA-Seq in 24 of 26 samples. SIFT and PolyPhen2 indicated both mutations would result in loss of function, and SNAP2 that they would be non-neutral. Amino acid substitutions projected no change of hydrophobicity and isoelectric point in PACRG, a change in both for RTTN; and a slight change in 3D structure for PACRG and RTTN. For *PACRG,* samples from additional individuals confirmed higher frequency of the heterozygous genotype in asthmatics, while the *RTTN* homozygous mutant phenotype was more prevalent in the asthmatic compared to non-asthmatic group.

64 **Discussion.** RNA-Seq was sensitive and specific for calling gene variants in this disease model.
65 Even moderate coverage (<10-20 cpm) yielded correct identification in 92% of samples,
66 suggesting RNA-Seq may be suitable to detect variants in low coverage samples. The impact of
67 amino acid alterations in PACRG and RTTN proteins are unknown at this point, but their role in
68 structure and function of cilia may warrant further investigation.
69
70
71

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 variants, and may be influenced by age and sex [18-20]. Several susceptibility gene 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].

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 central nervous system ventricles [39,40]. Motile cilia directionally propel cells or extracellular fluid through “metachronal wave” beating movements [41,42]. 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 [43], and protofilament ribbon proteins that are essential for sliding of adjacent microtubule doublets to generate ciliary movement [44]. Abnormalities in cilia are now appreciated as cause for the development of respiratory diseases, often through genetic variants associated with a loss of function affecting unique ciliary proteins [39]. Factors that affect beating, synchronization or orientation of motile cilia result in accumulation of mucus in airways [39]. Furthermore, hedgehog (HH) signaling is strongly linked to ciliary function, and many components of the HH signaling pathway localize to cilia [45,46].

Parkin co-regulated (*PACRG)* is a gene conserved across species [47], and shares a bi-directional promoter with parkin (*PARK2*) [48]. PACRG is affiliated with axonemal doublet microtubules, and contributes to the signaling pathway that controls dynein-driven microtubule sliding [47,49-51]. A single nucleotide variant (SNV) in *PACRG* was strongly associated with an increased risk of developing childhood asthma following early-life tobacco smoke exposure [52]. Rotatin *(RTTN)* is a cilium-associated protein [53] essential for assembly of centrosomes in non-motile and motile cilia [54]. Absence of RTTN, or presence of gene variants that disrupt the interaction of RTTN with SCL/TAL1 interrupting locus (*STIL*), abrogate proper ciliary development and function [54], and recessive mutations in *RTTN* are linked to abnormal primary ciliary development in humans [55]. Collectively, these findings incriminate that aspects of ciliary function may be impaired in asthma.

RNA-Seq is a promising approach for calling genetic 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, and whether specific variants were associated with the equine asthmatic phenotype.

METHODS

Animals and procedures

Animal procedures, sample collection and sample processing were as previously described [56]. Briefly, six horses with asthma in remission and seven horses without asthma that belong to the institutional research herd (mean ages of 15 and 12 years, respectively, p =0.352, unpaired t test) were placed in a dust-free indoor environment for 24 hours before exposure to dusty hay until respiratory impairment was apparent in asthmatic horses (range 1 to 3 days, average 2.2 days).

Non-asthmatic horses were exposed to dusty hay for 3 days. Physical examination, pulmonary function test, bronchoalveolar lavage and endoscopic bronchial biopsy were performed before and after exposure to the asthmatic challenge, with sequential biopsies obtained from the contralateral lung lobe. Samples from an additional four asthmatics and seven non-asthmatic horses were used for PCR-amplification of specific 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

Total RNA was extracted from endobronchial biopsies (Qiagen, Toronto, ON). RNA quality and concentration was determined with the Bioanalyzer RNA Nanochip (Agilent, Mississauga, ON) and capillary electrophoresis. RNA-Seq unstranded library preparation and sequencing were performed at The Centre for Applied Genomics (TCAG; Toronto, ON) using Illumina TruSeq RNA sample preparation and sequencing protocols (Illumina, San Diego, CA). For each sample, approximately 1 μg of non-degraded, high quality total RNA was enriched for poly-A RNA, fragmented into 200 to 300 bases, and converted to double stranded cDNA libraries. Final RNA libraries were quantified (KAPA Library Quantification kit, Kapa Biosystems, Wilmington, MA) prior to pooling and sequencing. Illumina flow cells were prepared and sequenced on an Illumina HiSeq 2500 instrument in 5 lanes following the manufacturer's instructions to generate paired-end reads of 100-bases.

Raw read quality was assessed using FastQC software version 0.10.1 [57] and aligned to the horse reference genome [58] (Ensembl v70) with STAR version 2.4 [59]. The STAR_pass2 alignment protocol was followed including these adaptations: 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.

Variant calling and filtering

Read processing, 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 (http://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 [60] and the *-T SplitNCigarReads*, *-rf ReassignOneMappingQuality*, *-RMQF 255*, *-RMQT 60* and *-U ALLOW_N_CIGAR_READS* options.

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 variants were processed with the variant filtration function of GATK software and the following options: *-window 35, -cluster 3, -filterName FS, -filter "FS > 30.0", -filterName QD* and *-filter "QD < 2.0"*.

Variants were analyzed individually in each of 26 samples (6 asthmatics and 7 non-asthmatics, before and after asthmatic challenge). Variants of interest were first identified based on presence in pre- and/or post-challenge samples from asthmatic horses. The STATS function in the SeqMule software [61] was used to identify consensus variants within groups with the *–c -vcf* options. Venn diagrams depicting variants shared between and within groups were constructed using the *-p –venn* options (Suppl. Figure 1). The Ensembl Variant Predictor (EVP) on-line tool [62] was used to further analyze and filter variants. In asthmatic horses, a total of 26,619 pre- and 24,527 post-challenge variants were identified, respectively, while the corresponding numbers were 28,909 and 28,451 for non-asthmatic horses. Approximately 30% of variants were novel and not previously described. The types of variants and their coding region effects are summarized in Suppl. Figures 2 and 3. For further variant selection using VEP the inclusion criteria were 1) missense mutation in protein-coding sequence; and 2) predicted to cause loss of protein function. The variant effect on protein function was analyzed with the Sorting Intolerant From Tolerant (SIFT) [63-67] tool and the threshold score was set at <0.01. Low confidence loss of function predictions and existing variants were excluded. Ten variants present in all asthmatic horses (before or/and after challenge) but not all non-asthmatic horses were identified. Of these, only two variants were expressed in all samples and were more prevalent in asthmatic compared to non-asthmatics. Effects of these two variants were then

194 investigated with Polymorphism Phenotyping (PolyPhen) 2 [68] and Screening for Non-
195 Acceptable Polymorphism (SNAP) 2 softwares [69-71].
196
197 **Sequence alignment and protein secondary structure prediction**
198 DNA sequence alignment and prediction of secondary structure was performed with Geneious
199 version 10.2.3. Default settings were used for both operations. Hydrophobicity and isoelectric
200 point (pI) were included for each amino acid of the protein sequence.
201
202 **PCR**
203 Primers for amplification of variant regions from bronchial cDNA were *PACRG* forward (5′
204 -CTC TGA ACC TCC GAA ACC GAC-3′) and reverse (5′-CTC CTG GGA TAA CTC ACC
205 ATT C-3′), and *RTTN* forward (5′-TCC TGA GTT GTA TCA AGA AGT G-3′) and reverse (5′-
206 CCA GCC TGC AAT TCC TTT CT-3′). A Taq polymerase PCR kit (Invitrogen, Mississauga,
207 ON) was used for PCR amplifications. Each reaction was performed in a 25 µL final volume,
208 including 5 µL of 10X PCR buffer, 0.2 mM dNTPs, 2 mM $MgSO_4$, 0.3 µM of each primer, 2 U
209 of Platinum Taq, and 5 µL (100 ng) of template DNA. PCR conditions for amplification were
210 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
211 *RTTN*, respectively, and 72 °C for 90 sec, followed by final elongation for 10 min at 72 °C.
212 Twenty µL of each PCR product was separated by electrophoresis in a 1% agarose gel stained
213 with SYBR Safe (Invitrogen). Amplicons of appropriate size were cut out and DNA extracted
214 and purified (QIAquick, Qiagen). Extracted and purified PCR products were Sanger sequenced
215 at Laboratory Services Division (Guelph, ON).
216
217 **RESULTS**
218 **RNA-Seq coverage**
219 RNA-coverage was visualized with Integrative Genomics Viewer (IGV) software**.** With the IGV
220 tool large-scale genomic data sets are visualized in real-time over a wide range of resolutions
221 [72]. Coverage (Fig. 1) was similar across horses and conditions except for slightly higher
222 coverage in some asthmatic horses following challenge.
223
224 **Variant calling and filtering**

The GATK variant calling and filtering workflow yielded 2823 and 1788 variants present specifically in the asthmatic group pre- and post-challenge, respectively (Suppl. Figure 1). Of these, 10 were missense mutations, coded for proteins and had SIFT scores <0.01. Variants in *PACRG* (Fig. 2A) and *RTTN* (Fig. 2B) had lowest prevalence in 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. 2A). 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. 2B).

Protein alignment

In PACRG, the G/A substitution resulted in replacement of valine (V) for methionine (M) at position 182 (Fig. 3A). PACRG sequence alignment of wild type (WT) and mutant proteins predicted changes from beta-strand to alpha-helix structure in the mutant protein a few amino acids distant from the site of substitution (182) at positions 187 and 188 (Fig. 3B). 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. 4A). Sequence alignment of WT and mutant proteins also indicated a change from alpha helix to beta strand structure near the site of substitution (bp 1807) at position 1816 (Fig. 4B). In addition, increased hydrophobicity and decreased pI were projected at the site of substitution (1807) in the mutant compared to the WT protein.

Expression of *PACRG* and *RTTN*

Expression of *PACRG* in counts-per-million (CPM) was overall higher than that of *RTTN* (Fig. 5). *PACRG* CPM ranged from ~30.7 to 66.3 in asthmatic horses (Fig. 5A) and ~25.1 to 65.6 in non-asthmatic horses (Fig. 5B), while expression of *RTTN* varied from ~11.5 to 24.9 and ~13.1 to 29.9 in asthmatic (Fig. 5C) and non-asthmatic (Fig. 5D) horses, respectively. Expression of *PACRG* overall was slightly higher in asthmatic than non-asthmatic horses.

Predicted effect of variant on protein function

The mutations detected in PACRG and RTTN were considered to cause loss of protein function when analyzed with SIFT and PolyPhen2 (Suppl. Fig. 4A and B). For PACRG, the mutation score was 0.993 with sensitivity of 0.70 and specificity of 0.97. For RTTN, the score was 0.979 with a sensitivity of 0.76 and a specificity of 0.96. Furthermore, for PACRG, V to M substitution at position 182 was predicted to affect protein function with a score of 64 at 80% expected accuracy (Table 1). For RTTN, R to W substitution at position 1807 would affect protein function with a score of 81 at 91% expected accuracy (Table 2).

Sanger sequencing

PACRG variants identified by RNA-Seq were confirmed with Sanger sequencing of PCR amplicons in asthmatic (Fig. 6A) and non-asthmatic (Fig. 6B) horses. In the asthmatic group, four horses were heterozygous [A/G] and two had homozygous-mutant alleles [A/A] (Fig. 6A). 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-mutant [A/A] (Fig. 6B). 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. 7). Four asthmatic horses were heterozygous [A/T] and two were homozygous-mutant [A/A] (Fig. 7A). 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 mutant [A/A] and none had the

287 homozygous-WT [G/G] genotype. Among non-asthmatic horses, six had the heterozygous
288 [A/G], five had the homozygous mutant [A/A] and three had the homozygous-WT [G/G]
289 genotype (Fig. 8A). *RTTN* alignment in asthmatic horses yielded five heterozygous [A/T], two
290 homozygous mutant [A/A], and three homozygous-WT [T/T] genotypes. In non-asthmatic
291 horses, four had the heterozygous [A/T], 10 had the homozygous-WT [T/T], and no horse had
292 the homozygous mutant [A/A] genotype (Fig. 8B).
293
294 **DISCUSSION**
295 The goal of this study was to assess the reliability of an adapted RNA-Seq variant calling
296 workflow compared to Sanger sequencing, and to identify genetic variants of potential interest in
297 asthmatic horses. Variant calling using RNA-Seq reads is recent practice, and reliability of
298 results is a function of sequencing platform, depth, quality, precision of read mapping and
299 appropriate variant calling and filtering methods. The reliability of identifying genetic variants
300 using RNA-Seq has been considered uncertain. In some reports RNA-Seq was considered useful
301 for identifying genetic variants [73,74] while in others differences between RNA and DNA
302 sequences were found with potentially frequent false results [75-77].
303 In this study we applied a modification of GATK best practices for variant calling with
304 RNA-Seq, and verified the results with Sanger sequencing. In 24 of 26 samples variants in
305 *PACRG* and *RTTN* were identified by both methods, while two horses' genotypes were
306 discordant by RNA-Seq with inconsistent genotypes before and after challenge. Sanger
307 sequencing confirmed one of the discordant genotypes, while the other could not be further
308 assessed.
309 We first visually assessed the overall genome coverage of RNA-Seq using IGV.
310 Although this method has limited precision, there was a similar expression pattern across all
311 horses, conditions and the entire genome, and expression appeared slightly higher in asthmatic
312 horses after challenge. This was expected since differential gene expression analysis previously
313 showed more prominent changes in asthmatic compared to non-asthmatic horses [56]. Two
314 candidate missense variants in the *PACRG* and *RTTN* coding sequence were identified after
315 variant filtering. SIFT was initially applied, followed by PolyPhen2 and SNAP2, to predict the
316 variant effect on protein function. SIFT uses phylogenetic data [63-67], while PolyPhen2 uses
317 structural information and multiple alignments [68] to predict whether or not a mutation may

cause loss of function. The two methods often yield similar results, but limited specificity suggests that results should be interpreted with caution [78]. SNAP2, on the other hand, uses evolutionary, structural, solvent-access and annotation information, as well as data from available homologs to predict whether a mutation is likely to have an effect or not on protein function [69-71]. While these three approaches can yield different results [79,80], inferences regarding PACRG and RTTN amino acid substitutions were consistent.

Presence of the mutation 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 expression and therefore lesser sequencing coverage, in particular in post-challenge samples, might have increased the likelihood of error in variant calling. Nonetheless, the vast majority of alleles were identified properly, suggesting that this workflow is suitable for variant calling in RNA-Seq at gene coverage in the 10 to 20 cpm range. Read counts were not filtered by common read counting algorithms, such as HTSeq, and may have included ambiguous reads warranting filtering for other purposes such as gene expression analysis.

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 mutations were considered to potentially cause loss of function and to have non-neutral effects (Tables 1 and 2). The PACRG mutation could impair or modify ability of the protein to bind interacting partners or form homodimers. For the RTTN mutation, 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 [81]. A change in the structural stability or binding affinity of the entire protein or the affected residue could impact ciliary structure and function. RTTN interacts with STIL and is essential for proper full-length centriole assembly [54]. The R1807W mutation in the carboxy-terminal region is not immediately at the suggested centrosome-targeting and STIL-binding site [54] but could nevertheless result in defective centrioles and hence cilium structure and function.

Variant sequence determination in 13 RNA-Seq and 11 additional samples showed that 80% of asthmatic animals were heterozygous and 20% were homozygous mutant for PACRG, and that no individual had the homozygous-WT genotype (G/G). Conversely, in non-asthmatic animals more than half were homozygous, whether WT or mutant (5 [A/A] and 3 [G/G]). In both groups the frequency of the homozygous-WT genotype was lower than either heterozygous genotype

For *RTTN*, 20% of asthmatic horses were homozygous mutant (A/A), 30% were homozygous WT (T/T) and 50% were heterozygous. Among non-asthmatics, none was homozygous mutant (A/A), while 71% of horses were homozygous WT (T/T) and 29% were heterozygous. Therefore, the mutation was present in 70% of asthmatic horses and in only 30% of non-asthmatic horses (heterozygous or homozygous mutant). Considering samples from non-asthmatic outnumbered those from asthmatic horses, there appears to be a clear trend for presence of the mutation in the latter group.

The PACRG protein associates with protofilaments [82] of the ciliary axoneme [47,83,84], has a role in ciliary morphogenesis and function [50] and is directly involved in ciliary motility through control of dynein-driven microtubule sliding [49]. 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 [51,85,86]. Impaired function or interaction of PACRG with its partners could weaken or impair ciliary stability and motility. We also suggest the possibility of homodimer formation by PACRG, which could explain the prevalence of the heterozygous genotype observed here and in pulmonary tuberculosis in humans [87]. A genome-wide interaction study also identified a *PACRG* SNP to be linked to an increased risk of childhood-onset asthma development following early-life exposure to tobacco smoke [52]. SNPs in *PACRG* also contributed to susceptibility to tuberculosis [88].

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 [89]. Interestingly, *PTCH1* is differentially expressed in asthmatic compared to non-asthmatic horses following challenge [56], linking *PACRG* and an asthmatic response to environmental agents with the HH pathway.

The exact nature and function of methionine in protein structure remains incompletely understood, and substitutions involving methionine has been associated with several diseases [90]. Both valine and methionine are hydrophobic residues grouped among the least polar amino acids [91]. Methionine is a sulfur-containing amino acid that is among the most hydrophobic residues and also easily oxidized if exposed [92]. 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 [93]. Methionine was overrepresented as a mutant residue in several mutations associated with decrease or loss of function [94], 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.

For *RTTN*, the R->W substitution altered the hydrophobicity and isoelectric point at position 1807. RTTN is a centrosome-associated protein first discovered for its role in axial rotation and left-right specification in the mouse embryo [53]. R <=> W substitutions were predicted to be the most problematic in the human genome [96], and R->W substitution is generally disfavored in all proteins types [93]. Overrepresentation of mutated arginine was a prominent feature among disease-causing mutations in a range of diseases [94]. *RTTN* appears to be highly conserved in exons but not introns (unpublished results), which suggests strong evolutionary pressure on the coding sequence and renders mutations in the coding sequence as particularly unusual.

CONCLUSIONS

Variants can be confidently called with RNA-Seq results as low as 10-20 CPM. Single point mutations in *PACRG* and *RTTN* were more prevalent in asthmatic compared to non-asthmatic horses. The heterozygous mutant genotype of *PACRG* was more prevalent among asthmatics while the homozygous-WT genotype of *RTTN* was more prevalent in non-asthmatics. Functional cilia are crucial for lung health, and mutations resulting in impaired function will have a negative impact. The significance of the substitutions in *PACRG* and *RTTN* remain to be determined but they are highly suggestive to affect ciliary function. Analysis of variant frequency at a population level and WT and mutant protein-protein interactions, and determining protein crystal structure, may be useful 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 JP, 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 JP. 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 JD, Lavoie JP. 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.
11. Klukowska-Rötzler J, Marti E, Lavoie J-P, Ainsworth DM, Gerber V, Zurbriggen A, Janda J. Expression of thymic stromal lymphopoietin in equine recurrent airway obstruction. Vet Immunol Immunopathol. 2012; 146:46-52.

12. Ainsworth DM, Grünig G, Matychak MB, Young J, Wagner B, Erb HN, Antczak DF. Recurrent airway obstruction (RAO) in horses is characterized by IFN-gamma and IL-8 production in bronchoalveolar lavage cells. Vet Immunol Immunopathol. 2003; 96:83-91.

13. Debrue M, Hamilton E, Joubert P, Lajoie-Kadoch S, Lavoie JP. Chronic exacerbation of equine heaves is associated with an increased expression of interleukin-17 mRNA in bronchoalveolar lavage cells. Vet Immunol Immunopathol. 2005; 105:25-31.

14. Ainsworth DM, Matychak M, Reyner CL, Erb HN, Young JC. Effects of in vitro exposure to hay dust on the gene expression of chemokines and cell-surface receptors in primary bronchial epithelial cell cultures established from horses with chronic recurrent airway obstruction. Am J Vet Res. 2009; 70:365-72.

15. Ainsworth DM, Wagner B, Erb HN, Young JC, Retallick DE. Effects of in vitro exposure to hay dust on expression of interleukin-17,-23,-8, and-1β and chemokine (CXC motif) ligand 2 by pulmonary mononuclear cells isolated from horses chronically affected with recurrent airway disease. Am J Vet Res. 2007; 68:1361-9.

16. Lavoie-Lamoureux A, Moran K, Beauchamp G, Mauel S, Steinbach F, Lefebvre-Lavoie J, Martin JG, Lavoie JP. IL-4 activates equine neutrophils and induces a mixed inflammatory cytokine expression profile with enhanced neutrophil chemotactic mediator release ex vivo. Am J Physiol Lung Cell Mol Physiol. 2010; 299:L472-82.

17. Padoan E, Ferraresso S, Pegolo S, Castagnaro M, Barnini C, Bargelloni L. Real time RT-PCR analysis of inflammatory mediator expression in recurrent airway obstruction-affected horses. Vet Immunol Immunopathol. 2013; 156:190-9.

18. Marti E, Gerber H, Essich G, Oulehla J, Lazary S. The genetic basis of equine allergic diseases 1. Chronic hypersensitivity bronchitis. Equine Vet J. 1991; 23:457-60.

19. Couëtil LL, Ward MP. Analysis of risk factors for recurrent airway obstruction in North American horses: 1,444 cases (1990-1999). J Am Vet Med Assoc. 2003; 223:1645-50.

20. Ramseyer A, Gaillard C, Burger D, Straub R, Jost U, Boog C, Marti E, Gerber V. Effects of genetic and environmental factors on chronic lower airway disease in horses. Journal of 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 SA, Wong KC, Illig T,

473 Vogelberg C, Weiland SK, von Mutius E, Abecasis GR, et al. Genetic variants regulating
474 ORMDL3 expression contribute to the risk of childhood asthma. Nature. 2007; 448:470-3.
475 22. Choudhry S, Taub M, Mei R, Rodriguez-Santana J, Rodriguez-Cintron W, Shriver MD, Ziv
476 E, Risch NJ, Burchard EG. Genome-wide screen for asthma in Puerto Ricans: evidence for
477 association with 5q23 region. Hum Genet. 2008; 123:455-68.
478 23. Galanter J, Choudhry S, Eng C, Nazario S, Rodríguez-Santana JR, Casal J, Torres-Palacios
479 A, Salas J, Chapela R, Watson HG, Meade K, LeNoir M, Rodríguez-Cintrón W, Avila PC,
480 Burchard EG. ORMDL3 gene is associated with asthma in three ethnically diverse populations.
481 Am J Respir Crit Care Med. 2008; 177:1194-200.
482 24. Tavendale R, Macgregor DF, Mukhopadhyay S, Palmer CN. A polymorphism controlling
483 ORMDL3 expression is associated with asthma that is poorly controlled by current medications.
484 J Allergy Clin Immunol. 2008; 121:860-3.
485 25. Madore A-M, Tremblay K, Hudson TJ, Laprise C. Replication of an association between
486 17q21 SNPs and asthma in a French-Canadian familial collection. Human genetics. 2008;
487 123:93-5.
488 26. Hancock DB, Romieu I, Shi M, Sienra-Monge JJ, Wu H, Chiu GY, Li H, del Rio-Navarro
489 BE, Willis-Owen SA, Willis-Owens SA, Weiss ST, Raby BA, Gao H, Eng C, Chapela R,
490 Burchard EG, Tang H, Sullivan PF, London SJ. Genome-wide association study implicates
491 chromosome 9q21.31 as a susceptibility locus for asthma in mexican children. PLoS Genet.
492 2009; 5:e1000623.
493 27. Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, Wilk JB,
494 Willis-Owen SA, Klanderman B, Lasky-Su J, Lazarus R, Murphy AJ, Soto-Quiros ME, Avila L,
495 Beaty T, Mathias RA, Ruczinski I, Barnes KC, Celedón JC, Cookson WO, et al. Genome-wide
496 association analysis identifies PDE4D as an asthma-susceptibility gene. Am J Hum Genet. 2009;
497 84:581-93.
498 28. Bisgaard H, Bønnelykke K, Sleiman PM, Brasholt M, Chawes B, Kreiner-Møller E, Stage M,
499 Kim C, Tavendale R, Baty F. Chromosome 17q21 gene variants are associated with asthma and
500 exacerbations but not atopy in early childhood. Am J Respir Crit Care Med. 2009; 179:179-85.
501 29. Wu H, Romieu I, Sienra-Monge JJ, Li H, del Rio-Navarro BE, London SJ. Genetic variation
502 in ORM1-like 3 (ORMDL3) and gasdermin-like (GSDML) and childhood asthma. Allergy.
503 2009; 64:629-35.

504 30. Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, Meyers DA, Bleecker ER. Genome-
505 wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. J Allergy
506 Clin Immunol. 2010; 125:328-335.e11.
507 31. Sleiman PM, Flory J, Imielinski M, Bradfield JP, Annaiah K, Willis-Owen SA, Wang K,
508 Rafaels NM, Michel S, Bonnelykke K, Zhang H, Kim CE, Frackelton EC, Glessner JT, Hou C,
509 Otieno FG, Santa E, Thomas K, Smith RM, Glaberson WR, et al. Variants of DENND1B
510 associated with asthma in children. N Engl J Med. 2010; 362:36-44.
511 32. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, von Mutius E, Farrall
512 M, Lathrop M, Cookson WOCM. A large-scale, consortium-based genomewide association
513 study of asthma. N Engl J Med. 2010; 363:1211-21.
514 33. Ferreira MA, Matheson MC, Duffy DL, Marks GB, Hui J, Le Souëf P, Danoy P, Baltic S,
515 Nyholt DR, Jenkins M. Identification of IL6R and chromosome 11q13. 5 as risk loci for asthma.
516 The Lancet. 2011; 378:1006-14.
517 34. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, Himes
518 BE, Levin AM, Mathias RA, Hancock DB, Baurley JW, Eng C, Stern DA, Celedón JC, Rafaels
519 N, Capurso D, Conti DV, Roth LA, Soto-Quiros M, Togias A, et al. Meta-analysis of genome-
520 wide association studies of asthma in ethnically diverse North American populations. Nat Genet.
521 2011; 43:887-92.
522 35. Swinburne JE, Bogle H, Klukowska-Rötzler J, Drögemüller M, Leeb T, Temperton E, Dolf
523 G, Gerber V. A whole-genome scan for recurrent airway obstruction in Warmblood sport horses
524 indicates two positional candidate regions. Mamm Genome. 2009; 20:504-15.
525 36. Klukowska-Rötzler J, Swinburne JE, Drögemüller C, Dolf G, Janda J, Leeb T, Gerber V. The
526 interleukin 4 receptor gene and its role in recurrent airway obstruction in Swiss Warmblood
527 horses. Anim Genet. 2012; 43:450-3.
528 37. Jost U, Klukowska-Rötzler J, Dolf G, Swinburne JE, Ramseyer A, Bugno M, Burger D, Blott
529 S, Gerber V. A region on equine chromosome 13 is linked to recurrent airway obstruction in
530 horses. Equine Vet J. 2007; 39:236-41.
531 38. Klukowska-Rötzler J, Gerber V, Leeb T. Association analysis of SNPs in the IL21R gene
532 with recurrent airway obstruction (RAO) in Swiss Warmblood horses. Anim Genet. 2012;
533 43:475-6.

534 39. Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. Nat Rev
535 Mol Cell Biol. 2017; 18:533-47.
536 40. Szymanska K, Johnson CA. The transition zone: an essential functional compartment of cilia.
537 Cilia. 2012; 1:10.
538 41. Eshel D, Priel Z. Characterization of metachronal wave of beating cilia on frog's palate
539 epithelium in tissue culture. J Physiol. 1987; 388:1-8.
540 42. Gheber L, Priel Z. On metachronism in ciliary systems: a model describing the dependence
541 of the metachronal wave properties on the intrinsic ciliary parameters. Cell Motil Cytoskeleton.
542 1990; 16:167-81.
543 43. Steffen W, Linck RW. Evidence for tektins in centrioles and axonemal microtubules. Proc
544 Natl Acad Sci U S A. 1988; 85:2643-7.
545 44. Linck RW, Norrander JM. Protofilament ribbon compartments of ciliary and flagellar
546 microtubules. Protist. 2003; 154:299.
547 45. Huangfu D, Liu A, Rakeman AS, Murcia NS, Niswander L, Anderson KV. Hedgehog
548 signalling in the mouse requires intraflagellar transport proteins. Nature. 2003; 426:83-7.
549 46. Goetz SC, Anderson KV. The primary cilium: a signalling centre during vertebrate
550 development. Nat Rev Genet. 2010; 11:331-44.
551 47. Thumberger T, Hagenlocher C, Tisler M, Beyer T, Tietze N, Schweickert A, Feistel K, Blum
552 M. Ciliary and non-ciliary expression and function of PACRG during vertebrate development.
553 Cilia. 2012; 1:13.
554 48. West AB, Lockhart PJ, O'Farell C, Farrer MJ. Identification of a novel gene linked to parkin
555 via a bi-directional promoter. J Mol Biol. 2003; 326:11-9.
556 49. Mizuno K, Dymek EE, Smith EF. Microtubule binding protein PACRG plays a role in
557 regulating specific ciliary dyneins during microtubule sliding. Cytoskeleton. 2016; 73:703-11.
558 50. Wilson GR, Wang HX, Egan GF, Robinson PJ, Delatycki MB, O'Bryan MK, Lockhart PJ.
559 Deletion of the Parkin co-regulated gene causes defects in ependymal ciliary motility and
560 hydrocephalus in the quakingviable mutant mouse. Hum Mol Genet. 2010; 19:1593-602.
561 51. Ikeda T. Parkin-co-regulated gene (PACRG) product interacts with tubulin and microtubules.
562 FEBS Lett. 2008; 582:1413-8.
563 52. Scholtens S, Postma DS, Moffatt MF, Panasevich S, Granell R, Henderson AJ, Melén E,
564 Nyberg F, Pershagen G, Jarvis D, Ramasamy A, Wjst M, Svanes C, Bouzigon E, Demenais F,

565 Kauffmann F, Siroux V, von Mutius E, Ege MJ, Braun-Fahrländer C, et al. Novel childhood
566 asthma genes interact with in utero and early-life tobacco smoke exposure(). J Allergy Clin
567 Immunol. 2014; 133:885-8.
568 53. Faisst AM, Alvarez-Bolado G, Treichel D, Gruss P. Rotatin is a novel gene required for axial
569 rotation and left-right specification in mouse embryos. Mech Dev. 2002; 113:15-28.
570 54. Chen HY, Wu CT, Tang CC, Lin YN, Wang WJ, Tang TK. Human microcephaly protein
571 RTTN interacts with STIL and is required to build full-length centrioles. Nat Commun. 2017;
572 8:247.
573 55. Kheradmand Kia S, Verbeek E, Engelen E, Schot R, Poot RA, de Coo IF, Lequin MH,
574 Poulton CJ, Pourfarzad F, Grosveld FG, Brehm A, de Wit MC, Oegema R, Dobyns WB,
575 Verheijen FW, Mancini GM. RTTN mutations link primary cilia function to organization of the
576 human cerebral cortex. Am J Hum Genet. 2012; 91:533-40.
577 56. Tessier L, Côté O, Clark ME, Viel L, Diaz-Méndez A, Anders S, Bienzle D. Impaired
578 response of the bronchial epithelium to inflammation characterizes severe equine asthma. BMC
579 Genomics. 2017; 18:708.
580 57. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence
581 Data. 2015. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. December 2, 2015.
582 58. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear TL, Adelson DL,
583 Bailey E, Bellone RR. Genome sequence, comparative analysis, and population genetics of the
584 domestic horse. Science. 2009; 326:865-7.
585 59. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
586 Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29:15-21.
587 60. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,
588 Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce
589 framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297-
590 303.
591 61. Guo Y, Ding X, Shen Y, Lyon GJ, Wang K. SeqMule: automated pipeline for analysis of
592 human exome/genome sequencing data. Sci Rep. 2015; 5:srep14283.
593 62. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the
594 consequences of genomic variants with the Ensembl API and SNP Effect Predictor.
595 Bioinformatics. 2010; 26:2069-70.

596 63. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res. 2001;
597 11:863-74.
598 64. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein
599 function. Genome Res. 2002; 12:436-46.
600 65. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic
601 Acids Res. 2003; 31:3812-4.
602 66. Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function.
603 Annu Rev Genomics Hum Genet. 2006; 7:61-80.
604 67. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on
605 protein function using the SIFT algorithm. Nat Protoc. 2009; 4:1073-81.
606 68. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense
607 mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013; Chapter 7:Unit7.20.
608 69. Hecht M, Bromberg Y, Rost B. News from the protein mutability landscape. J Mol Biol.
609 2013; 425:3937-48.
610 70. Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function.
611 Nucleic Acids Res. 2007; 35:3823-35.
612 71. Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence variants.
613 BMC Genomics. 2015; 16 Suppl 8:S1.
614 72. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP.
615 Integrative Genomics Viewer. Nat Biotechnol. 2011; 29:24-6.
616 73. Quinn EM, Cormican P, Kenny EM, Hill M, Anney R, Gill M, Corvin AP, Morris DW.
617 Development of strategies for SNP detection in RNA-seq data: application to lymphoblastoid
618 cell lines and evaluation using 1000 Genomes data. PLoS One. 2013; 8:e58815.
619 74. Piskol R, Ramaswami G, Li JB. Reliable identification of genomic variants from RNA-seq
620 data. Am J Hum Genet. 2013; 93:641-51.
621 75. Sheng Q, Zhao S, Li CI, Shyr Y, Guo Y. Practicability of detecting somatic point mutation
622 from RNA high throughput sequencing data. Genomics. 2016; 107:163-9.
623 76. Guo Y, Zhao S, Sheng Q, Samuels DC, Shyr Y. The discrepancy among single nucleotide
624 variants detected by DNA and RNA high throughput sequencing data. BMC Genomics. 2017;
625 18:690.

626 77. Li X, Howard TD, Moore WC, Ampleford EJ, Li H, Busse WW, Calhoun WJ, Castro M,
627 Chung KF, Erzurum SC, Fitzpatrick AM, Gaston B, Israel E, Jarjour NN, Teague WG, Wenzel
628 SE, Peters SP, Hawkins GA, Bleecker ER, Meyers DA. Importance of hedgehog interacting
629 protein and other lung function genes in asthma. J Allergy Clin Immunol. 2011; 127:1457-65.
630 78. Flanagan SE, Patch AM, Ellard S. Using SIFT and PolyPhen to predict loss-of-function and
631 gain-of-function mutations. Genet Test Mol Biomarkers. 2010; 14:533-7.
632 79. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X. Comparison and integration
633 of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing
634 studies. Hum Mol Genet. 2015; 24:2125-37.
635 80. Thusberg J, Vihinen M. Pathogenic or not? And if so, then how? Studying the effects of
636 missense mutations using bioinformatics methods. Hum Mutat. 2009; 30:703-14.
637 81. Betts MJ, Russell RB. Amino Acid Properties and Consequences of Substitutions. In:
638 Bioinformatics for Geneticists. John Wiley & Sons, Ltd; 2003. p. 289-316.
639 82. Ikeda K, Ikeda T, Morikawa K, Kamiya R. Axonemal localization of Chlamydomonas
640 PACRG, a homologue of the human Parkin-coregulated gene product. Cell Motil Cytoskeleton.
641 2007; 64:814-21.
642 83. Lechtreck KF, Luro S, Awata J, Witman GB. HA-tagging of putative flagellar proteins in
643 Chlamydomonas reinhardtii identifies a novel protein of intraflagellar transport complex B. Cell
644 Motil Cytoskeleton. 2009; 66:469-82.
645 84. Dawe HR, Farr H, Portman N, Shaw MK, Gull K. The Parkin co-regulated gene product,
646 PACRG, is an evolutionarily conserved axonemal protein that functions in outer-doublet
647 microtubule morphogenesis. J Cell Sci. 2005; 118:5421-30.
648 85. Li Y, Xiao X, Ji X, Liu B, Amos CI. RNA-seq analysis of lung adenocarcinomas reveals
649 different gene expression profiles between smoking and nonsmoking patients. Tumour Biol.
650 2015; 36:8993-9003.
651 86. Imai Y, Soda M, Murakami T, Shoji M, Abe K, Takahashi R. A product of the human gene
652 adjacent to parkin is a component of Lewy bodies and suppresses Pael receptor-induced cell
653 death. J Biol Chem. 2003; 278:51901-10.
654 87. Udina IG, Kordicheva SIu, Pospelov LE, Malenko AF, Gergert VIa, Pospelov AL,
655 Matrashkin AG, Kyzyl-Ool MM, Nachin AA, Zhivotovskiĭ LA. [Study of the polymorphic

656 markers--the PARK2 and PACRG genes due to the incidence of pulmonary tuberculosis in two
657 districts of the Republic of Tyva]. Probl Tuberk Bolezn Legk. 2007:27-9.
658 88. Bragina EY, Tiys ES, Rudko AA, Ivanisenko VA, Freidin MB. Novel tuberculosis
659 susceptibility candidate genes revealed by the reconstruction and analysis of associative
660 networks. Infect Genet Evol. 2016; 46:118-23.
661 89. Gavino C, Richard S. Patched1 haploinsufficiency impairs ependymal cilia function of the
662 quaking viable mice, leading to fatal hydrocephalus. Mol Cell Neurosci. 2011; 47:100-7.
663 90. Valley CC, Cembran A, Perlmutter JD, Lewis AK, Labello NP, Gao J, Sachs JN. The
664 methionine-aromatic motif plays a unique role in stabilizing protein structure. Journal of
665 Biological Chemistry. 2012; 287:34979-91.
666 91. Wolfenden R. Experimental measures of amino acid hydrophobicity and the topology of
667 transmembrane and globular proteins. J Gen Physiol. 2007; 129:357-62.
668 92. Brosnan JT, Brosnan ME. The sulfur-containing amino acids: an overview. J Nutr. 2006;
669 136:1636S-40S.
670 93. Betts MJ, Russell RB. Amino acid properties and consequences of substitutions.
671 Bioinformatics for geneticists. 2003; 317:289.
672 94. Khan S, Vihinen M. Spectrum of disease-causing mutations in protein secondary structures.
673 BMC Struct Biol. 2007; 7:56.
674 95. Kazemi-Esfarjani P, Beitel LK, Trifiro M, Kaufman M, Rennie P, Sheppard P, Matusik R,
675 Pinsky L. Substitution of valine-865 by methionine or leucine in the human androgen receptor
676 causes complete or partial androgen insensitivity, respectively with distinct androgen receptor
677 phenotypes. Molecular Endocrinology. 1993; 7:37-46.
678 96. Majewski J, Ott J. Amino acid substitutions in the human genome: evolutionary implications
679 of single nucleotide polymorphisms. Gene. 2003; 305:167-73.

LEGENDS

Figure 1. RNA-Seq coverage visualized with IGV software and illustrated by a heat map graph. White indicates <10 counts, and gradually darkening blue indicates raw counts from 10 (light blue) to a maximum of 5000 (dark blue). Sample origin from asthmatic and non-asthmatic horses pre- and post-challenge is indicated on the y-axis. Chromosomal location (1-31, X) is indicated on the top, and annotations from Refseq (Refseq genes) and Ensembl (EqCab2.70.gtf) on the bottom. Coverage is similar across different samples, with slightly higher coverage in some asthmatic horses post-challenge.

Figure 2. 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 variant (lower) are included with gene sequence at the bottom. Location of the variant is indicated in blue for NCBI and Ensembl databases.

Figure 3. 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 4. 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 5. 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.

712 **Figure 6**. Comparison of GATK variant calls and Sanger sequencing results for *PACRG* in
713 asthmatic (A) and non-asthmatic (B) horses. For both groups, the bar graph indicates the IGV
714 count for each allele (A-red, G-yellow), horse and condition. Below the bar graph is the GATK
715 variant call, the electropherogram of the Sanger sequence, and agreement. (A) Four asthmatic
716 horses (1, 2, 5, 6) had heterozygous alleles [A/G] and two (3 and 4) were homozygous for the
717 mutant allele [A/A]. (B) In non-asthmatic horses, one horse (6) had heterozygous alleles [A/G],
718 three horses (1, 2 and 4) were homozygous for the wild-type allele [G/G] and three horses (3, 5
719 and 7) were homozygous for the mutant allele [A/A]. All genotypes were consistent across
720 horses and methods. DNA was not available for non-asthmatic horse 1.
721
722 **Figure 7**. Comparison of GATK variant calls and Sanger sequencing results for *RTTN* in
723 asthmatic (A) and non-asthmatic (B) horses. Details as in Fig. 6. (A) Four asthmatic horses (3-6)
724 had heterozygous alleles [A/T] and two (1 and 2) were homozygous for the mutant allele [A/A].
725 Genotypes were consistent across horses and methods. (B) In non-asthmatic horses, one (5) had
726 heterozygous [A/T] alleles, four horses had homozygous wild type [T/T] alleles, and two horses
727 (1 and 4) were inconsistently identified as homozygous wild type and heterozygous in different
728 samples. Sanger sequencing confirmed the genotype of horse 4 as heterozygous. DNA was not
729 available for non-asthmatic horse 1.
730
731 **Figure 8**. Alignment of PACRG (A) and RTTN (B) Sanger sequences for 10 asthmatic and 14
732 non-asthmatic horses with the reference genome. (A) For PACRG, 8 asthmatic horses (80%)
733 were heterozygous [A/G], 2 (20%) were homozygous mutant [A/A] and none was homozygous
734 wild-type [G/G]. Six non-asthmatic horses (43%) were heterozygous [A/G], 5 (36%) were
735 homozygous mutant [A/A] and 3 (21%) were homozygous wild type [G/G]. (B) For RTTN, in
736 the asthmatic group, there were 5 (50%) heterozygous [A/T], 2 (20%) homozygous mutant [A/A]
737 and 3 (30%) homozygous wild type [T/T] genotypes. In the non-asthmatic group, 4 (29%) were
738 heterozygous [A/G], 10 (71%) were homozygous wild type [T/T], and none had the homozygous
739 mutant [A/A] genotype.
740

Figure 1

RNA-Seq coverage visualized with IGV software and illustrated by a heat map graph.

White indicates <10 counts, and gradually darkening blue indicates raw counts from 10 (light blue) to a maximum of 5000 (dark blue). Sample origin from asthmatic and non-asthmatic horses pre- and post-challenge is indicated on the y-axis. Chromosomal location (1-31, X) is indicated on the top, and annotations from Refseq (Refseq genes) and Ensembl (EqCab2.70.gtf) on the bottom. Coverage is similar across different samples, with slightly higher coverage in some asthmatic horses post-challenge.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3491v1 | CC BY 4.0 Open Access | rec: 27 Dec 2017, publ: 28 Dec 2017

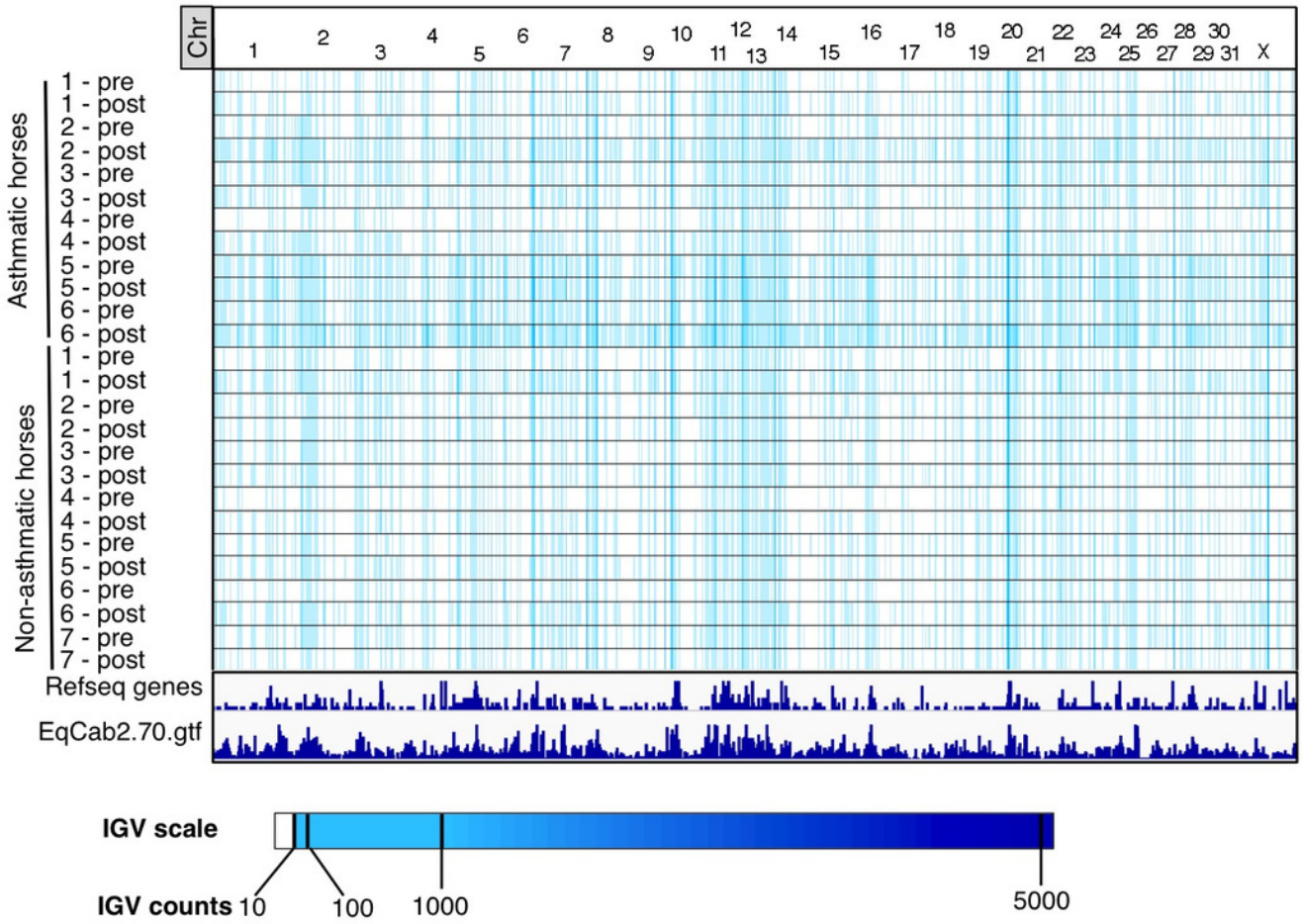


Figure 2

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 variant (lower) are included with gene sequence at the bottom. Location of the variant is indicated in blue for NCBI and Ensembl databases.

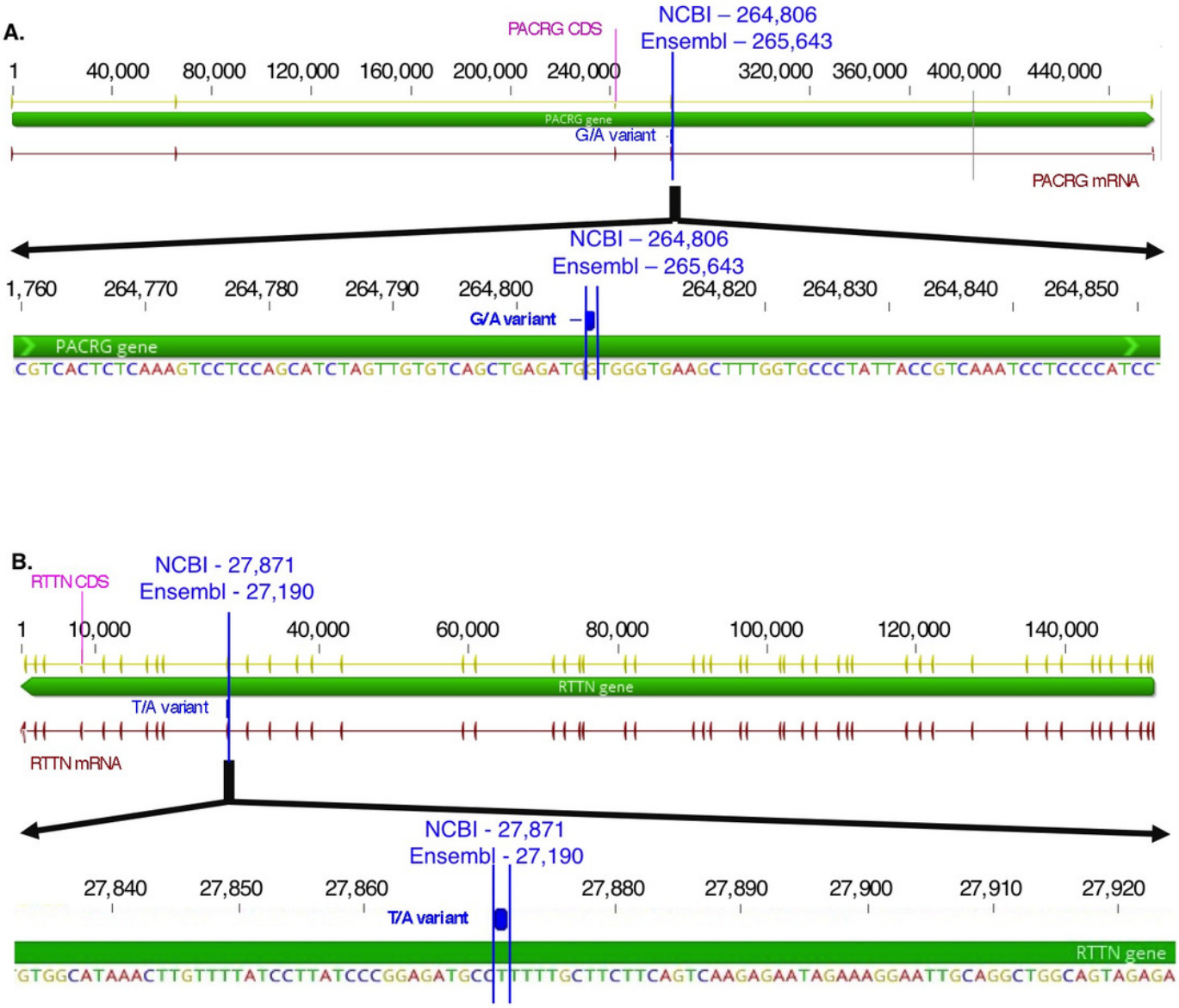


Figure 3

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).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3491v1 | CC BY 4.0 Open Access | rec: 27 Dec 2017, publ: 28 Dec 2017

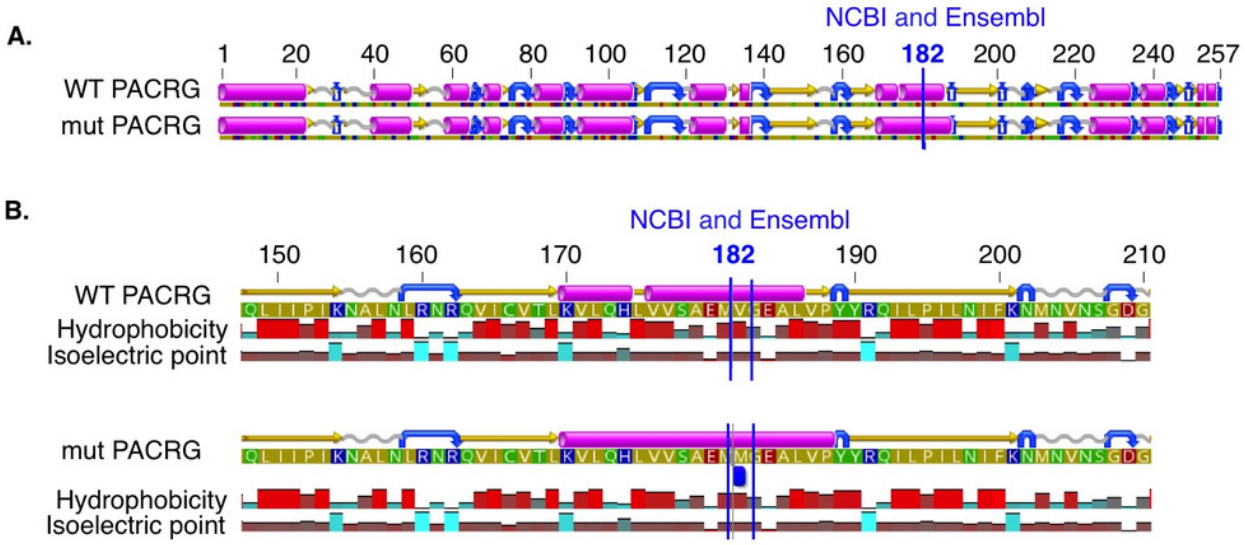


Figure 4

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).

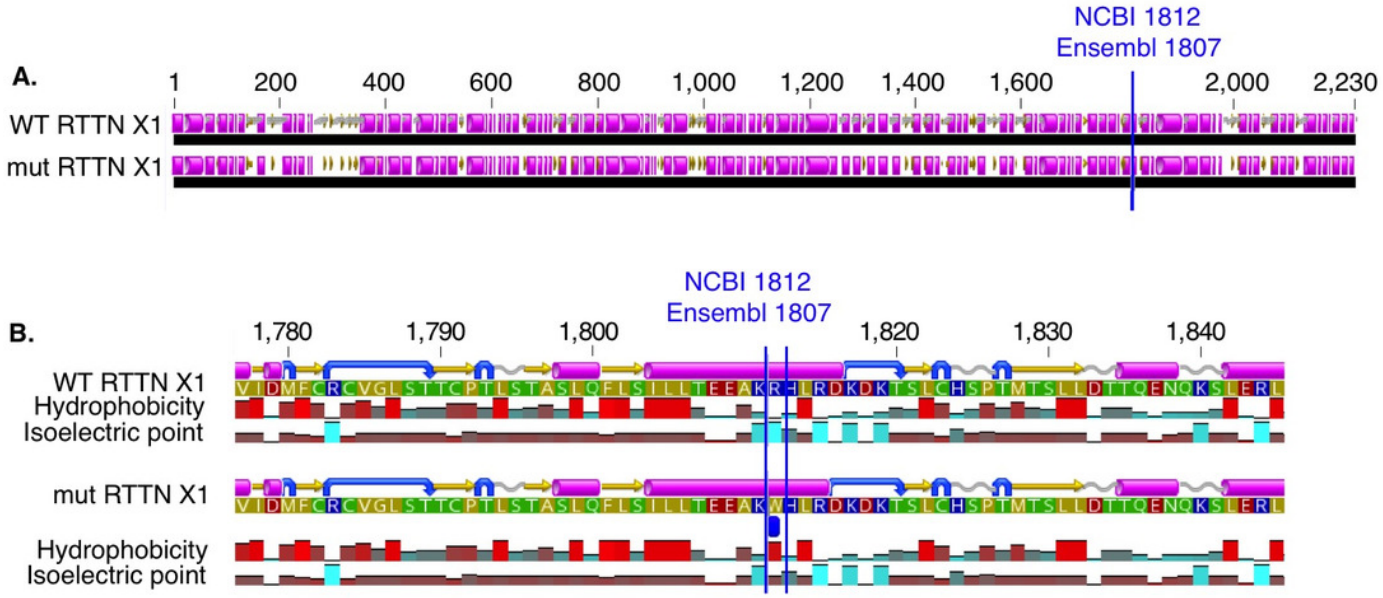


Figure 5

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.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3491v1 | CC BY 4.0 Open Access | rec: 27 Dec 2017, publ: 28 Dec 2017

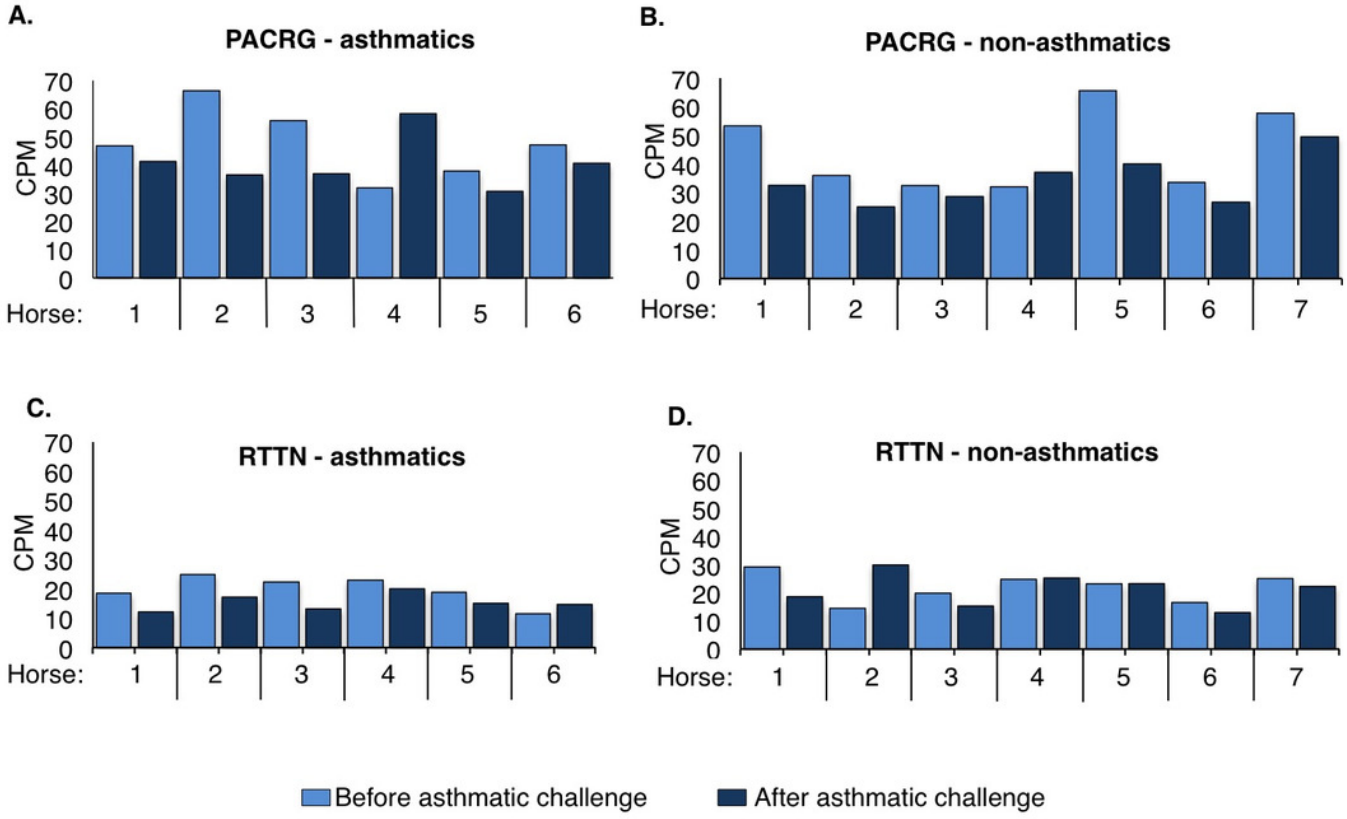


Figure 6

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.3491v1 | CC BY 4.0 Open Access | rec: 27 Dec 2017, publ: 28 Dec 2017

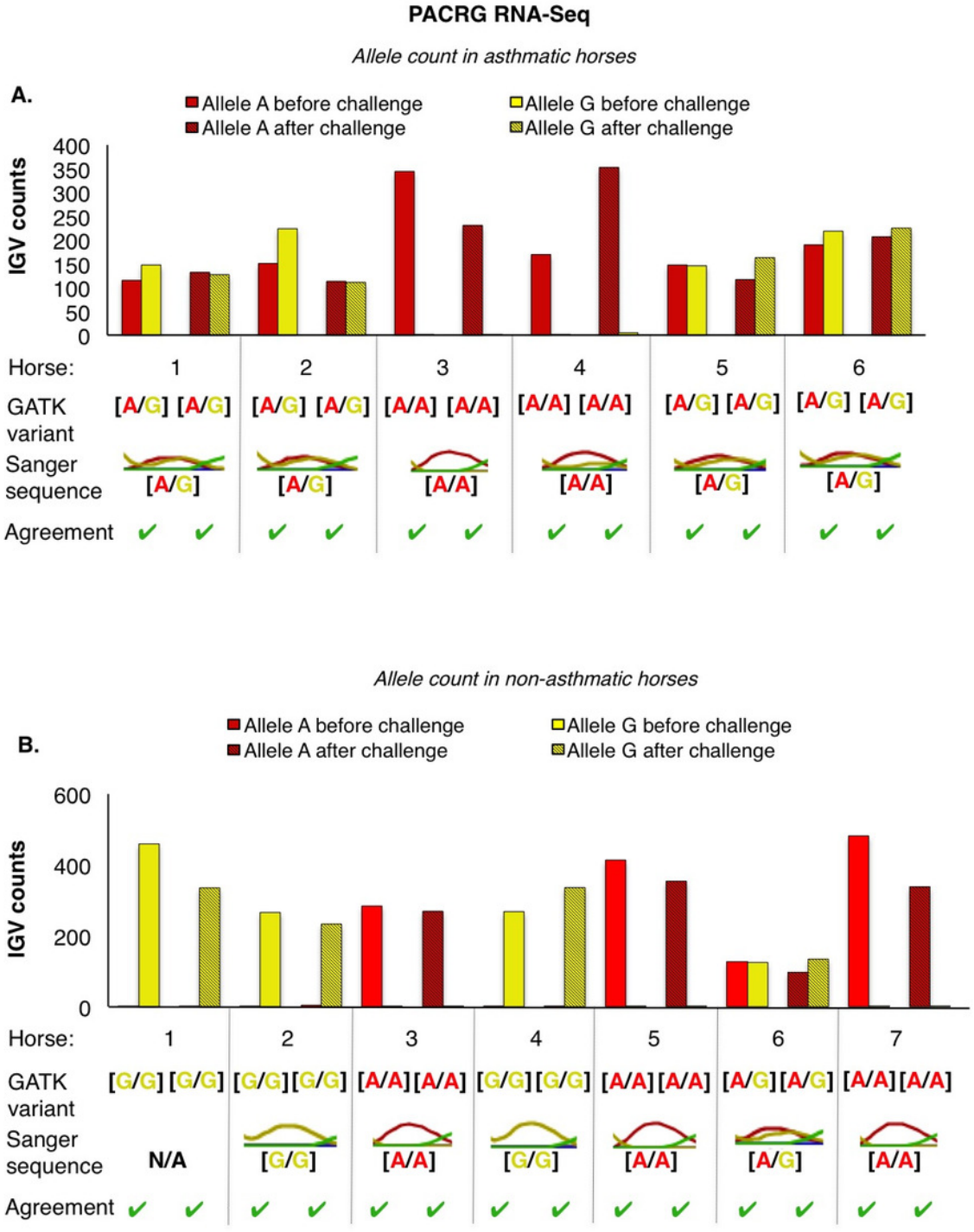


Figure 7

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.

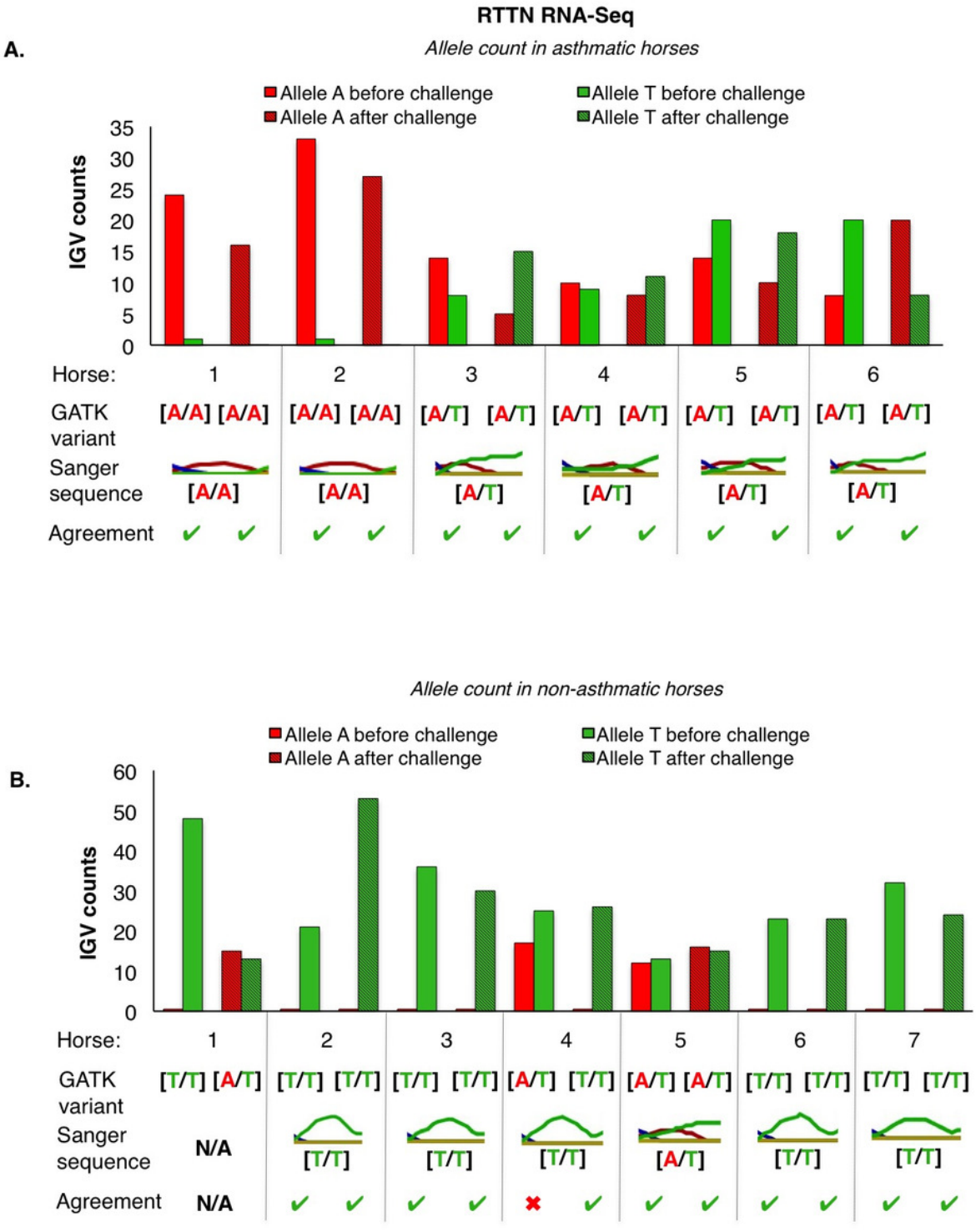
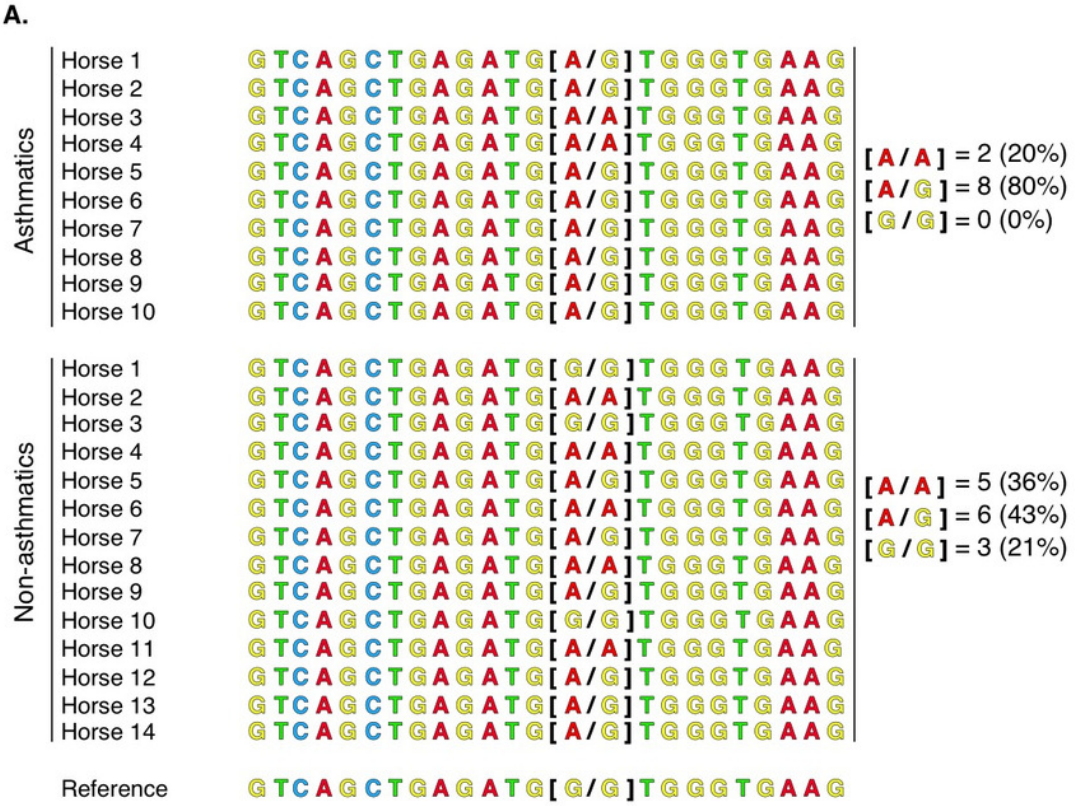


Figure 8

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.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3491v1 | CC BY 4.0 Open Access | rec: 27 Dec 2017, publ: 28 Dec 2017


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

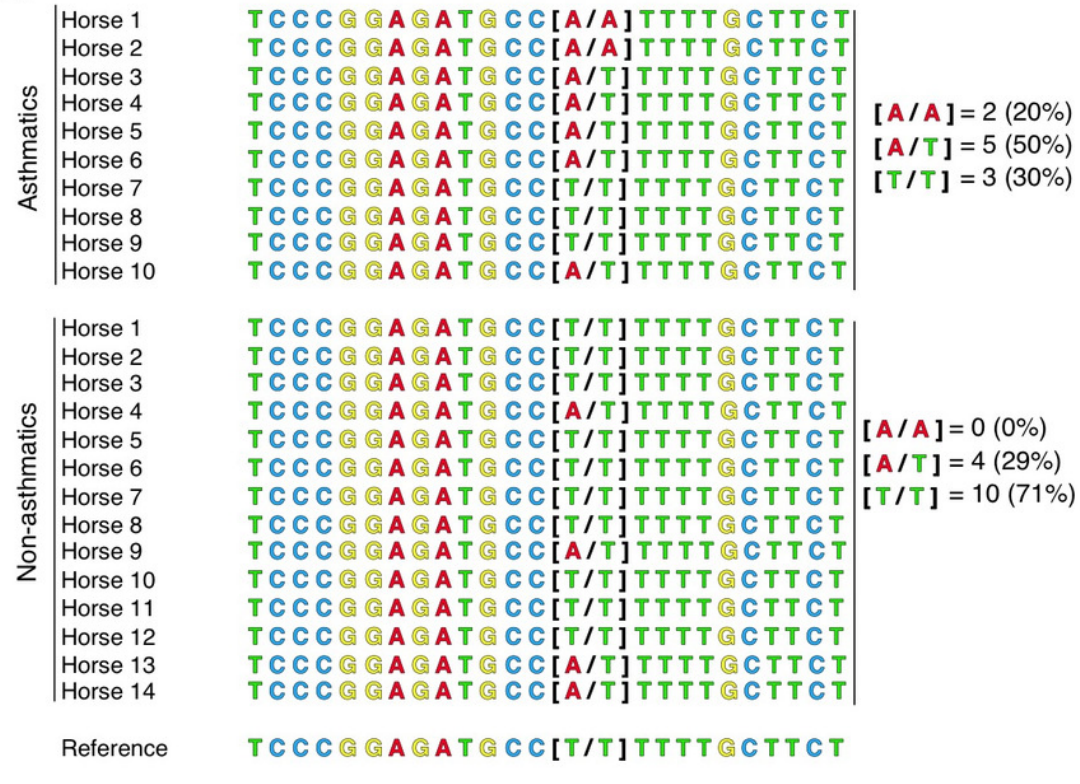

B.
Asthmatics
Horse 1 TCCCGGAGATGCC[A/A]TTTTGCTTCT
Horse 2 TCCCGGAGATGCC[A/A]TTTTGCTTCT
Horse 3 TCCCGGAGATGCC[A/T]TTTTGCTTCT
Horse 4 TCCCGGAGATGCC[A/T]TTTTGCTTCT
Horse 5 TCCCGGAGATGCC[A/T]TTTTGCTTCT
Horse 6 TCCCGGAGATGCC[A/T]TTTTGCTTCT
Horse 7 TCCCGGAGATGCC[T/T]TTTTGCTTCT
Horse 8 TCCCGGAGATGCC[T/T]TTTTGCTTCT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

Table 1(on next page)

Table 1. Predicted effect of substitution at position 182 in PACRG

Table 1. Predicted effect of substitution at position 182 in PACRG

Variant	Outcome	Score	Accuracy
V182A	neutral	-1	53%
V182R	effect	85	91%
V182N	effect	75	85%
V182D	effect	81	91%
V182C	effect	34	66%
V182Q	effect	78	85%
V182E	effect	75	85%
V182G	effect	78	85%
V182H	effect	82	91%
V182I	neutral	-59	78%
V182L	effect	65	80%
V182K	effect	86	91%
V182M	effect	64	80%
V182F	effect	66	80%
V182P	effect	83	91%
V182S	effect	32	66%
V182T	effect	51	75%
V182W	effect	84	91%
V182Y	effect	79	85%
V182V	neutral	-85	93%

Table 2(on next page)

Table 2. Predicted effect of substitution at position 1807 for RTTN

Table 2. Predicted effect of substitution at position 1807 for RTTN

Variant	Outcome	Score	Accuracy
R1807A	effect	44	71%
R1807R	neutral	-99	97%
R1807N	effect	54	75%
R1807D	effect	80	91%
R1807C	effect	64	80%
R1807Q	effect	34	66%
R1807E	effect	67	80%
R1807G	effect	55	75%
R1807H	effect	43	71%
R1807I	effect	62	80%
R1807L	effect	49	71%
R1807K	effect	21	63%
R1807M	effect	60	80%
R1807F	effect	74	85%
R1807P	effect	84	91%
R1807S	effect	45	71%
R1807T	effect	50	75%
R1807W	effect	81	91%
R1807Y	effect	67	80%
R1807V	effect	66	80%