UC Davis

UC Davis Previously Published Works

Title

Defining Trends in Global Gene Expression in Arabian Horses with Cerebellar Abiotrophy

Permalink

https://escholarship.org/uc/item/2tk3r44q

Journal

The Cerebellum, 16(2)

ISSN

1473-4222

Authors

Scott, EY
Penedo, MCT
Murray, JD
et al.

Publication Date

2017-04-01

DOI

10.1007/s12311-016-0823-8

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Cerebellum
DOI 10.1007/s12311-016-0823-8

ORIGINAL PAPER

Defining Trends in Global Gene Expression in Arabian Horses with Cerebellar Abiotrophy

E. Y. Scott[1] · **M. C. T. Penedo**[2,3] · **J. D. Murray**[1,3] · **C. J. Finno**[3]

© Springer Science+Business Media New York 2016

Abstract Equine cerebellar abiotrophy (CA) is a hereditary neurodegenerative disease that affects the Purkinje neurons of the cerebellum and causes ataxia in Arabian foals. Signs of CA are typically first recognized either at birth to any time up to 6 months of age. CA is inherited as an autosomal recessive trait and is associated with a single nucleotide polymorphism (SNP) on equine chromosome 2 (13074277G>A), located in the fourth exon of *TOE1* and in proximity to *MUTYH* on the antisense strand. We hypothesize that unraveling the functional consequences of the CA SNP using RNA-seq will elucidate the molecular pathways underlying the CA phenotype. RNA-seq (100 bp PE strand-specific) was performed in cerebellar tissue from four CA-affected and five age-matched unaffected horses. Three pipelines for differential gene expression (DE) analysis were used (Tophat2/Cuffdiff2, Kallisto/EdgeR, and Kallisto/Sleuth) with 151 significant DE genes identified by all three pipelines in CA-affected horses. *TOE1* (Log_2(foldchange) = 0.92, p = 0.66) and *MUTYH* (Log_2(foldchange) = 1.13, p = 0.66) were not differentially expressed. Among the major pathways that were differentially expressed, genes associated with calcium homeostasis and specifically expressed in Purkinje neurons, *CALB1* (Log_2(foldchange) = −1.7, $p < 0.01$) and *CA8* (Log_2(foldchange) = −0.97, $p < 0.01$), were significantly down-regulated, confirming loss of Purkinje neurons. There was also a significant up-regulation of markers for microglial phagocytosis, *TYROBP* (Log_2(foldchange) = 1.99, $p < 0.01$) and *TREM2* (Log_2(foldchange) = 2.02, $p < 0.01$). These findings reaffirm a loss of Purkinje neurons in CA-affected horses along with a potential secondary loss of granular neurons and activation of microglial cells.

Keywords Cerebellar abiotrophy · Horse · Calcium · Microglial activation · RNA-seq

Genetics: Section Editor - Antoni Matilla-Dueñas

Electronic supplementary material The online version of this article (doi:10.1007/s12311-016-0823-8) contains supplementary material, which is available to authorized users.

✉ J. D. Murray
JDMurray@ucdavis.edu

✉ C. J. Finno
cjfinno@ucdavis.edu

E. Y. Scott
eyscott@ucdavis.edu

M. C. T. Penedo
mctorrespenedo@ucdavis.edu

[1] Department of Animal Science, University of California, Davis, USA

[2] Veterinary Genetics Laboratory, School of Veterinary Medicine, University of California, Davis, USA

[3] Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, USA

Introduction

Cerebellar abiotrophy (CA) is a neurodegenerative disease often resulting from the loss of Purkinje neurons in the cerebellum. Among the abiotrophies seen in domestic animals, CA is one of the most common [1] and often involves an autosomal recessive mode of inheritance. Mutations associated with CA vary between species, with an ultimate diagnosis of CA arising from an incomplete loss of Purkinje and, in some species, secondary loss of granular neurons [2–4]. In horses, CA results in varying levels of ataxia with exaggerated forelimb movement, difficulty rising from a recumbent position, head tremors and lack of

Published online: 05 October 2016

menace response [5]. These signs are consistent with the histological hallmark of apoptosis of Purkinje neurons and subsequent disorganization of the three-layer cerebellar cortex structure [6]. The loss of Purkinje neurons in CA has also been coupled with incomplete loss of granular neurons and a proliferation of Bergmann glia [7]. The equine CA phenotype varies in degree of severity as well as time of onset, which can be between a few days to approximately 6 months after birth [8]; no studies connecting the number of Purkinje neurons lost to severity or onset have been done. Unlike CA in several other species, a candidate single nucleotide polymorphism (SNP) on ECA2:13074277(G>A) (CA SNP) has been associated with the CA phenotype [9]; however, molecular mechanisms underlying the defect and the functionality of this SNP have yet to be determined. CA occurs almost exclusively in Arabian horses [10], with an average breed-wide carrier frequency of 20 % based on CA SNP screening of nearly 11,000 horses by the Veterinary Genetics Laboratory, UC Davis.

Horses with CA have inherited the CA SNP in an autosomal recessive manner [10]. The CA SNP causes an amino acid change, Arg > His, in exon 4 (R95H) of Target of Egr1 (*TOE1*) [9] (Fig. 1). Although both amino acids have similar characteristics, the change is in a highly conserved region of *TOE1* and could have a functional impact on *TOE1*. Alternatively, the overlapping transcriptional regions of *TOE1* and the promoter of the gene encoding MUTY homolog (*MUTYH)*, which is antisense to *TOE1*, suggest a potential interference of transcriptional regulation of *MUTYH* by the CA SNP. Neither *MUTYH* nor *TOE1* is Purkinje neuron specific; however, their regulation or potential functional importance during development of Purkinje neurons prevents them from being eliminated as candidate genes.

TOE1 is a mRNA deadenylase (adenine-specific exonuclease) that is also involved in nuclear Cajal body [11] and cytoplasmic Processing body (P-body) [12] formation. Cajal bodies are compartments within the nucleus thought to be associated with pre-mRNA processing [13]. P-bodies are RNA processing bodies known to contain proteins involved in mRNA metabolism and translational repression [14] and are known to respond to synaptic activity in mammalian neurons [15]. Despite Cajal bodies being associated with neurodegenerative disorders [16] and the likely presence of P-bodies in Purkinje neurons, *TOE1* was previously considered to have quite low levels of expression in the equine cerebellum [9], thus making *MUTYH* a more likely candidate gene for CA.

MUTYH is a DNA glycosylase involved in DNA repair in the nucleus and mitochondria in response to oxidative damage, specifically 8-oxoguanine [17]. The distribution of the nuclear and mitochondrial isoforms of *MUTYH* differs quite markedly during neurodevelopment in the rat, with nuclear isoforms being more prevalent in the embryo and mitochondrial isoforms in the neonate [18], which suggests a role of *MUTYH* in nuclear and then mitochondrial DNA repair during development. *MUTYH* has also been associated with two pathways leading to apoptosis of the cell, depending on whether the oxidative damage is primarily affecting the nucleus or the mitochondria [17]. Dysregulation of *MUTYH* has been linked to several neurodegenerative disorders [19], making it a promising candidate gene for CA.

The predicted *MUTYH* gene structure is incompletely annotated in the horse, lacking start and stop codons and a vast number of isoforms seen in humans [20], making targeted analysis of *MUTYH* difficult. A pilot study involving RNA sequencing (RNA-seq) to detect global differentially expressed (DE) genes associated with the observed Purkinje neuron loss in CA-affected horses was done. We hypothesized that the expression of *MUTYH*, and corresponding downstream pathways, would be decreased in CA-affected individuals. In addition, gene networks and their association with the CA phenotype were analyzed using unsupervised clustering (WGNCA) [21] and supervised clustering (ROAST) [22]. By defining global gene expression associated with the CA phenotype, we sought to uncover potential mechanisms to determine the functional consequences of the CA SNP and variable expressivity of the phenotype.

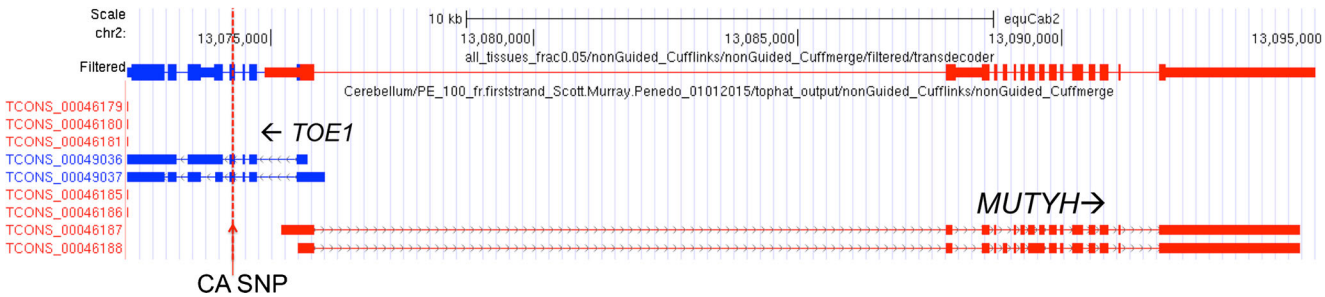

Fig. 1 UCSC Genome Browser depiction of *MUTYH* and *TOE1* genes in equine cerebellum. All tracks are from a publically available equine transcriptome (https://github.com/drtamermansour/horse_trans.git). The first track is a merged annotation of the region with transcriptomes from 8 different tissues. The *lower set* of tracks are cerebellar tissue-specific annotations of the region, where *TOE1* is in *blue*, *MUTYH* in *red*, and the CA SNP is represented by the *dotted red line and red arrow*

Method

Samples

Cerebella transcriptomes of five unaffected and four CA-affected horses were used for RNA-seq. Unaffected horses included a 3-month-old Paso Fino colt, a 6-month-old Thoroughbred gelding, a 1-year-old Quarter horse gelding, a 4-year-old Thoroughbred mare, and a 29-year-old Thoroughbred mare. The unaffected phenotype of these samples was confirmed via histologic assessment by board-certified pathologists at the UC Davis School of Veterinary Medicine, and all unaffected samples were not homozygous for the CA SNP [9]. The Arabian CA-affected horses included a 4-month-old colt, an 8-month-old mare, a 9-month-old colt, and a 6-year-old gelding. All CA-affected samples were homozygous for the CA SNP [9], with histologic confirmation of the disease at necropsy confirming Purkinje neuron loss. Cerebella samples were collected within an hour of euthanasia for all individuals. Samples were harvested by different veterinarians, with the instruction to extract sagittal cerebellar cortex slices from the cerebellum and either snap freeze or store in RNAlater (Thermofisher Scientific, Grand Island, NY, USA) at −80 °C. Sections of cerebella were also stored in formalin for histopathology examination.

RNA Extraction

Total RNA was extracted using TRIzol reagent (Thermofisher, Wilmington, DE, USA) and an electric homogenizer (TissueMiser, Fisher Scientific, Pittsburg, PA, USA). The RNA was concentrated, eluted, and treated with DNase (Promega, Madison, WI, USA) on Zymo RNA extraction columns (Zymo Research, Irvine, CA, USA) according to manufacturer's instructions. Quantification and quality of RNA were monitored on Bio-Rad Experion HiSens Chips (Bio-Rad, Hercules, CA, USA), with a quality threshold RNA quality indicator (RQI) of over 8.

cDNA Synthesis

cDNA was synthesized using the SuperScript II reverse transcriptase kit (Invitrogen, Carlsbad, CA, USA). One microgram of RNA was incubated with 75 ng of Oligo dT_{15} [23] primer and 1 μl dNTPs (10 mM each) at 65 °C for 5 min. First-Strand Buffer (5X), DTT (0.1 M), and RNAse Out (40 units) were added to tubes which were then incubated at 42 °C for 1 min. One microliter (200 units) of Superscript II was then added, and the reaction was incubated at 42 °C for 1 h followed by 70 °C for 15 min.

RNA-Seq Library Preparation

Total RNA was depleted of rRNA using the Epicentre Ribo-Zero Gold Kit for Human/Mouse/Rat samples, and 1 μg of this RNA was used to generate strand-specific PE100bp libraries with BIOO NEXTflex Rapid Directional RNA-Seq Kit (Bioo-Scientific, Austin, TX). Library qualities were analyzed with the Bioanalyzer 2100 (Agilent, Santa Clara, CA) and quantified on a Qubit instrument (LifeTechnologies, Carlsbad, CA). The libraries were then pooled in equimolar ratios, quantified by qPCR with a Kapa Library Quant kit (Kapa, Cape Town, South Africa), and sequenced on two lanes of an Illumina HiSeq 2500 run (Illumina, San Diego, CA).

RNA-seq Read Trimming and Mapping

All reads were trimmed with Sickle [24] with a Phred quality threshold of 30. Reads were then mapped either directly to EquCab2.0, 2007 (http://www.ncbi.nlm.nih.gov/genome/145) with Tophat2 [25] or with Kallisto [26] to an equine ENSEMBL, 2007 reference transcriptome (http://www.ensembl.org/info/data/ftp/index.html), with 30 bootstraps per sample. Only concordantly mapped reads were kept for further analysis.

Differentially Expressed Genes Analysis

Three DE analysis pipeline analyses were performed: Tophat2/Cuffdiff2 [27], Kallisto/EdgeR [28], and Kallisto/Sleuth (https://github.com/pachterlab/sleuth). Sleuth and EdgeR analyses were performed on the Kallisto output and Cuffdiff2 on the Tophat2 output (Supplementary Fig. 1, SF1). The input for Sleuth and EdgeR was normalized transcripts and counts per million, respectively. Cuffdiff2 normalizes input as fragments per kilobase per millions of reads (FPKM). All analysis programs are parametric; however, they have different statistical tests for differential gene expression. Sleuth uses the Wald test (https://github.com/pachterlab/sleuth), EdgeR uses the exact test [28], and Cuffdiff2 uses the T-test [27]. Only DE genes found in agreement between all three pipelines were further analyzed. Cuffdiff2 results were visualized with the CummeRbund package in R. EdgeR and Sleuth had built in visualization tools in R [29].

qPCR

All qPCR were performed in triplicate on an ABI 7500 Fast thermocycler (Applied Biosystems, Foster City, CA, USA) and analyzed with the ABI 7500 Software v2.0.1 (Applied Biosystems, Foster City, CA, USA). qPCR was performed on the same samples used for RNA-seq as validation of RNA-seq results. Fast SYBR Green MasterMix (Applied

Biosystems, Foster City, CA, USA) with 5 μM of forward and reverse primer (Supplementary Table S1) and approximately 50 ng of cDNA per reaction were used. The thermocycler stages consisted of 1 cycle of 95 °C for 20 s for enzyme activation and 40 cycles of 95 °C for 15 s and 60 °C for 30 s. Standard curves were generated from pooled cDNA from all samples (equimolar ratios) in concentrations of 200, 100, 50, 25, 12.5, and 0 ng/reaction. Using the standard curve, nanogram (ng) quantities of gene expression were averaged across the triplicates and Log_2(fold change) was calculated relative to the control and to the housekeeping gene *HPRT1*. *HPRT1* was selected as the best housekeeping gene over *SYNAP* because it showed consistent expression in all samples. Based on the small samples size, the nonparametric Mann-Whitney *U* test was used to calculate significance of Log_2(fold change) in CA-affected individuals and a Bonferroni corrected *p* value ($p = 0.0125$) was used to determine significance.

Unsupervised Clustering of RNA-seq Data with WGCNA

Unsupervised clustering was performed with WGCNA (weighted gene correlation network analysis) in R [21]. Hierarchically clustered genes were integrated into larger gene networks based on a correlation matrix depicting how correlated the clustered modules were to each other. WGCNA accounts for hub genes and expands gene networks to allow for detection of several genes that may be co-expressed on varying scales [21]. Normalized variance-stabilized read counts from EdgeR were used as input for WGCNA. The soft power threshold used was 15, with an unsigned network and height thresholds of 0.25 and 0.65. Using weighted correlation values between groups of genes, which were summarized as eigenvalues, more expansive gene networks were created and their correlation to the CA-affected phenotype was quantified. All figures were made within the WGCNA program.

Gene Set Testing of RNA-seq Data with ROAST

ROAST is a rotation gene set test that can identify whether a set of genes is differentially expressed between treatment groups [22]. It is especially relevant to this gene set because of its resilience against several experimental designs, including experiments with a small number of replicates. ROAST was used in conjunction with EdgeR and its output of normalized read counts (CPM). The gene sets provided were Purkinje neuron and granular neuron markers from Kuhn et al. [30] as well as Allen Brain Atlas derived markers from Kirsch et al. [31], with 236 and 1684 transcripts, respectively. All Purkinje neuron markers had visually confirmed localization to Purkinje neurons according to the Allen Brain Atlas ISH images. The calcium homeostasis markers (176 transcripts) were acquired from Bettencourt et al. [32] as well as from the Panther family of genes "calcium mediated signaling" [33]. Resulting MA plots were constructed using ROAST within the EdgeR package.

Panther Pathway Analysis

Gene lists consisting of ENSEMBL gene IDs from the 151 DE genes, Purkinje module, and Module 5 from the WGCNA analysis were used for the Panther Statistical Overrepresentation test [34]. Panther's output of biological processes and gene ontology (GO) molecular functions with *p* values below the Bonferroni-corrected threshold (5 % experiment-wide) for each of the three gene sets were further examined.

Results

RNA-seq Overall Mapping Statistics

Averaged across all samples, there were 34 million reads per sample with an average Phred quality score of 34. After Sickle trimming, the overall mapping rate to EquCab2.0, 2007 (http://www.ncbi.nlm.nih.gov/genome/145) was 91 %, 86 % of which was concordant.

Clustering of Treatment Groups

The five unaffected and four CA-affected horses clustered by treatment groups in a dendrogram (Fig. 2). The dendrogram

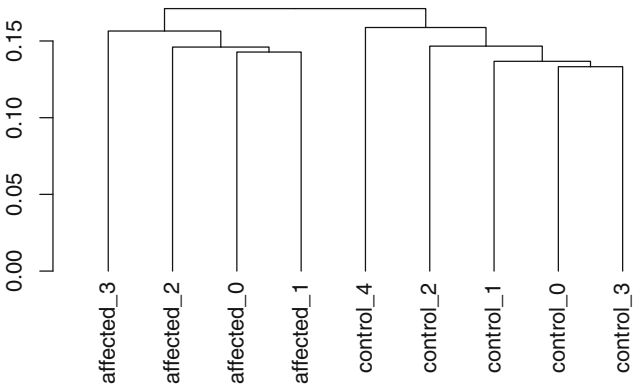

Fig. 2 Clustering of samples into their respective treatment groups. In this figure, and subsequent figures, the affected samples are named as such: affected_0 is the 4-month-old colt, affected_1 is the 7–8-month-old mare, affected_2 is the 9-month-old colt, and affected_3 is the 6-year-old gelding. For the unaffected, control_0 is the 3-month-old Paso Fino colt, control_1 is the 6-month-old Thoroughbred gelding, control_2 is the 1- year-old Quarter horse gelding, control_3 is the 4-year-old Thoroughbred mare, and control_4 is the 29-year-old Thoroughbred mare. The control and CA-affected horses cluster into their control or CA-affected treatment groups based upon a dendrogram using the Jensen-Shannon distance between the distribution of reads. Figure made with CummeRbund (25)

clusters based on the Jensen-Shannon distance between the distributions of variance-stabilized reads for each sample.

Differential Gene Analysis

There was a difference in the number of DE genes detected between the EdgeR [28], Sleuth and Cuffdiff2 [27] pipelines (Supplementary Table S2). The 151 DE genes (Supplementary Table S3) that were detected and in agreement between all three pipelines were further analyzed using heatmaps depicting Log_2(foldchange) relative to the control horses (Fig. 3a, b). The heatmaps also have the genes hierarchically clustered based on the read distribution of the gene, with clusters suggesting co-expression of the genes within the cluster. Clusters that demonstrated clear and contrasting differences between controls and CA-affected horses were outlined, and genes involved in calcium homeostasis or associated with glial activation were marked. Among the up-regulated DE genes, we detected a cluster of genes associated with microglia activation and, more specifically, with genes

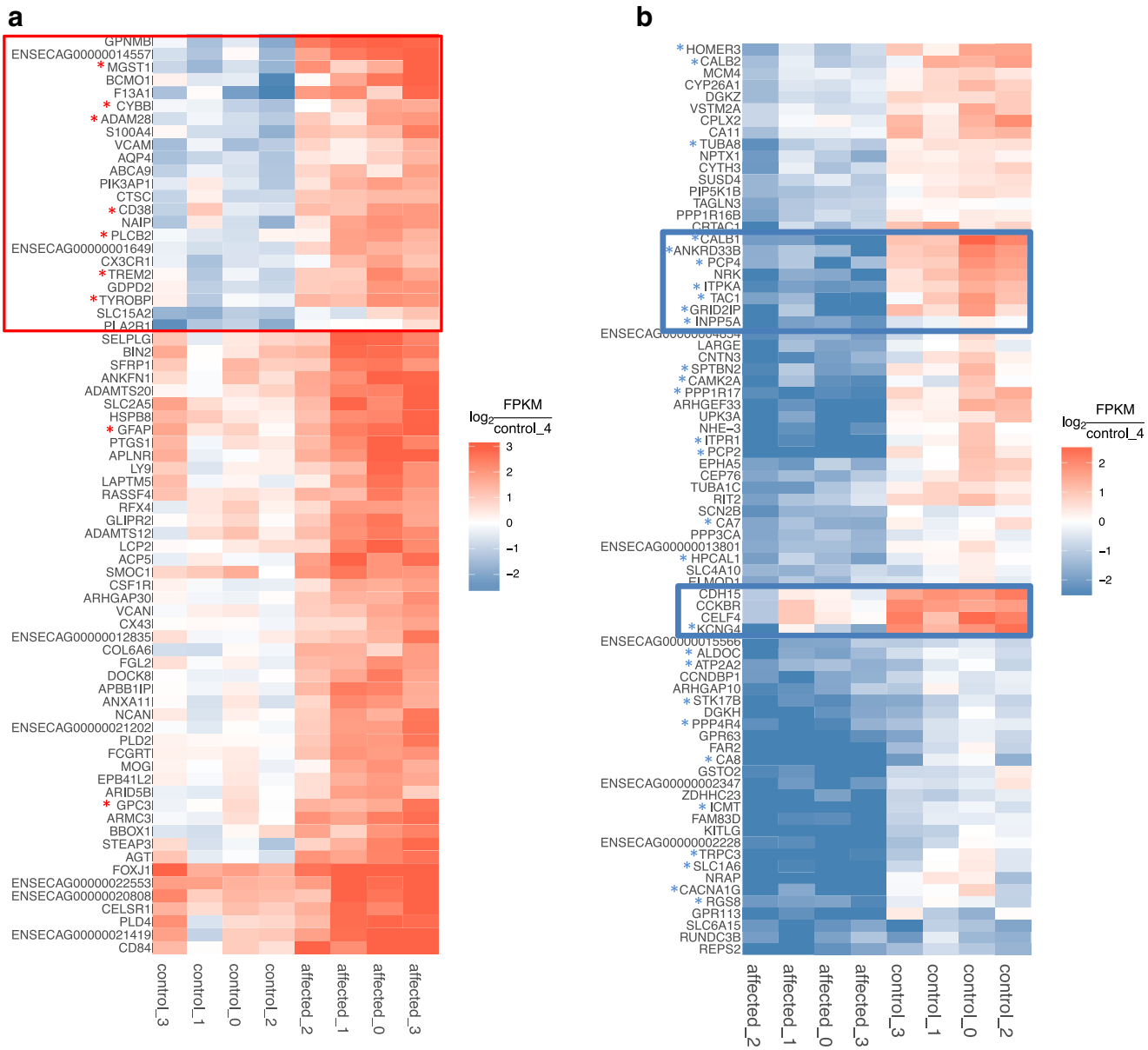

Fig. 3 Hierarchically clustered heatmaps showing gene expression in CA-affected horses (n = 4) by Log2(foldchange) relative to the unaffected (n = 4) separated into upregulated DE genes (**a**) and downregulated genes (**b**). Clusters that were further annotated with gene names are outlined in *blue* for down-regulated genes and *red* for up-regulated genes. The upregulated genes had a cluster associating with microglia, indicated with a *red asterisk*, and the downregulated genes had clusters with calcium homeostasis associated genes, indicated with *blue asterisks*. The 29-year-old control was removed from this analysis to obtain even treatment groups and appropriate Log2(foldchange) analysis

associated with microglia phagocytosis of apoptotic cells (Triggering Receptor Expressed On Myeloid Cells 2, *TREM2*, and TYRO Protein Tyrosine Kinase Binding Protein, *TYROBP*) (Fig. 3a, red asterisk). Among the down-regulated DE genes, we identified a cluster of genes involved in calcium homeostasis pathways (Fig. 3b, teal asterisk) in both Purkinje and granular neurons. The gene clusters resulting from hierarchical clustering contained genes with biological relevance to each other and suggested pathways associated with microglia activation to be up-regulated and pathways associated with calcium homeostasis to be down-regulated in CA-affected horses.

Unsupervised Clustering of RNA-seq Data with WGCNA

Unsupervised clustering was used to parse out cell-type specific gene expression. Gene networks were formulated based on the correlation of gene modules to one another. Upon post-clustering annotation of the larger gene networks with predefined markers for Purkinje neurons and granular neurons [30, 31, 35, 36], the turquoise gene network block, labeled as Purkinje neuron cluster, was enriched for Purkinje neuron gene markers, while the coral gene network-labeled granular neuron cluster was enriched mainly of granular neuron markers (Fig. 4a, Supplementary Table S4). The height threshold was then increased to 0.65 to obtain larger gene modules that might be correlated with the CA-phenotype (Modules 1 through 5), of which only Module 5 was significantly correlated with the CA phenotype (Fig. 4b, Supplementary Table S5). Module 5 contains 10 Purkinje neuron gene markers (among them Calbindin 1 (*CALB1*), Like-Glycosyltransferase (*LARGE*) and Polymerase (DNA-Directed), Epsilon 4 (*POLE4)*), and 11 calcium homeostasis markers, including Carbonic Anhydrase VIII (*CA8)*, Inositol-Trisphosphate 3-Kinase A (*ITPKA)*, and Calcium/Calmodulin-Dependent Protein Kinase II Alpha (*CAMK2A*). *MUTYH* and *TOE1*, however, were not in Module 5, but rather clustered in Module 1.

Gene Set Testing of RNAseq Data with ROAST

Given pre-defined markers [30–33, 35, 36], ROAST tests how gene sets behave, in terms of Log_2(foldchange), relative to the unaffected group, and determines the proportion of genes within a gene set that are dysregulated. Granular neuron versus Purkinje neuron marker gene expressions were compared and 40 % (FDR = 0.0002) of Purkinje neuron markers were down-regulated and 5 % considered up-regulated in the CA-affected group (Table 1), with a corresponding, however, non-significant down-regulation of granular neuron markers as well (21 %, FDR = 0.056). Among the calcium homeostasis markers, 40 % (FDR = 0.0004) were down-regulated in CA-affected horses (Fig. 5 and Table 1), of which 17 % of the calcium markers are also Purkinje neuron markers. Detailed expression analysis for the granular, Purkinje, and calcium homeostasis-associated genes can be found in Supplementary Tables 6, 7, and 8, respectively. Gene expression patterns for all genes can be found in Supplementary Table 9. In summary, this analysis demonstrated that markers for Purkinje neurons and calcium homeostasis were both down-regulated in CA-affected horses.

qPCR

Complete annotations of *MUTYH* and *TOE1* were obtained from the RNA-seq analysis (Fig. 1) and allowed for better primer design and more accurate qPCR analysis. qPCR confirmed results from RNA-seq and cluster analyses. In addition to *MUTYH* and *TOE1*, two genes related to calcium homeostasis (*CALB1* and *CA8*) were also chosen for qPCR, as both are markers for Purkinje neurons and involved in other cerebellar ataxias [32]. *MUTYH* and *TOE1* were not differentially expressed between control and CA-affected individuals ($p = 0.66$ for both genes); however, both *CALB1* and *CA8* were significantly down-regulated in CA-affected individuals by a Log_2 (fold change) of −1.7 ($p = 0.0043$) and −0.97 ($p = 0.0087$), respectively (Fig. 6).

Panther Pathway Analysis

The main biological processes over-represented in the 151 DE genes include cellular calcium ion homeostasis (fold enrichment >5, $p = 3.28E-2$) and overall homeostatic processes (fold enrichment >5, $p = 1.69E-2$) (Table 2, Supplementary Table S10). For the Purkinje module, which is a smaller component of the Module 5 from the WGCNA analysis, DNA repair (fold enrichment = 2.05, $p = 8.08E-5$) and mRNA processing (fold enrichment = 1.99, $p = 9.16E-8$) were the top two overrepresented biological processes (Table 2, Supplementary Table S11). For the larger Module 5, overrepresented biological processes included general cellular metabolic processes (fold enrichment = 1.43, $p = 1.69E-9$) and, more specifically, overrepresented GO molecular functions included ribonucleoside binding (fold enrichment = 1.52, $p = 4.10E-2$) (Table 2, Supplementary Table S12).

Discussion

Cerebellar abiotrophy is a neurologic defect that occurs in several species [37–40] and is characterized by the end result of loss of Purkinje neurons. The mechanisms resulting in Purkinje neuron loss within and across species are not well understood. Identification of a putative mutation in Arabian horses allows for a more targeted investigation of global gene expression associated with the CA phenotype in horses. Our

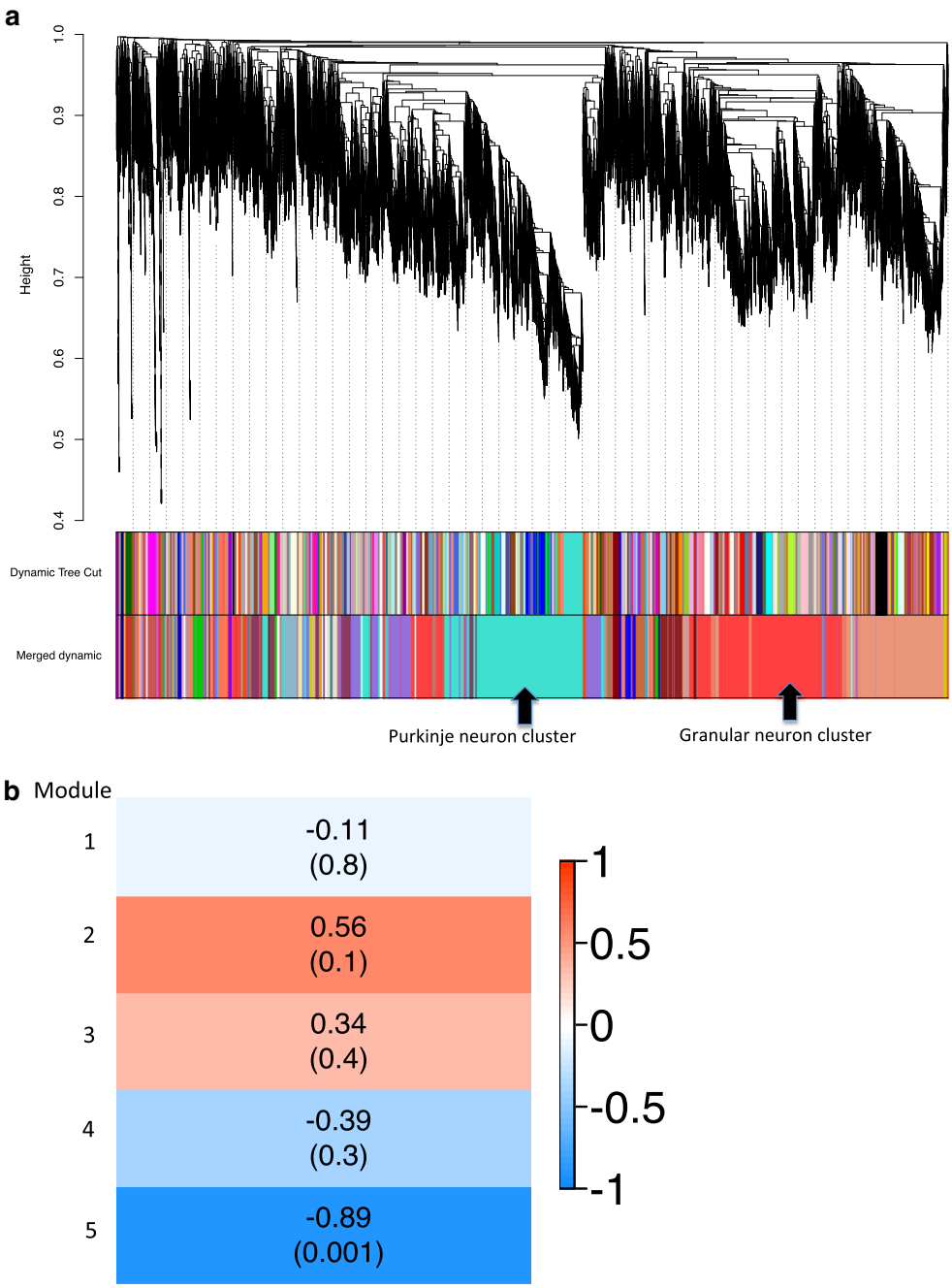

Fig. 4 WGCNA unsupervised clustering of gene networks in the five unaffected and four CA-affected horses. Initially, a topological overlay matrix (19) evaluates correlations between gene modules, with the gene modules (with a height threshold of 0.25) corresponding to the different colored blocks under the gene network dendrogram (**a**). The correlation (and corresponding *p* value in parentheses) of the gene modules, with a height threshold of 0.65, to the unaffected or CA-affected treatment group identify. Module 5 as the only significant gene module correlated to the CA-affected phenotype (**b**). The more negative (and *blue*) the module, the more correlated it is with the CA-affected phenotype

results provided evidence of up-regulated microglia pathways and down-regulated Purkinje neuron-associated calcium homeostasis markers in CA-affected horses.

Table 1 ROAST quantification regarding proportion of gene markers downregulated along with their corresponding false discovery rate (FDR) value

Gene set	Proportion downregulated	FDR
Granular	0.20962	0.05438
Purkinje	0.4025424	0.00046
Ca^{2+} genes	0.3988439	0.00041

Among the differentially expressed (DE) genes detected by our three DE pipelines, there were two distinct biological groups of genes demonstrating contrasting differences between unaffected and CA-affected individuals. In CA-affected horses, genes associated with microglia activation were up-regulated and contained *TREM2* and *TYROBP*, which are often affiliated with the microglia genetic signature and have a proposed involvement in phagocytosis of apoptotic cells [41]. Microglia are the major immune cells in the central nervous system

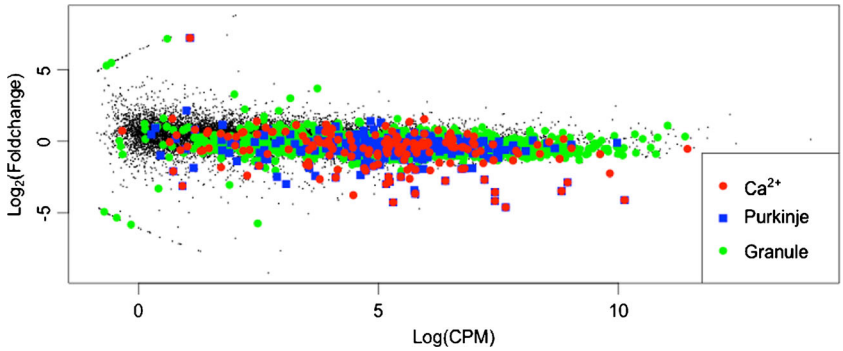

Fig. 5 ROAST analysis on the five unaffected and four CA-affected demonstrating how predefined gene sets correlate with control or CA-affected phenotype. MA plots identify granule cell marker (*green*), Purkinje neuron markers (*blue squares*), and calcium homeostasis markers (*red*) with respect to their Log2(foldchange) relative to controls using the normalized read count output from EdgeR

(CNS) and are responsible for adjusting synapses, removing cellular debris and mediating the inflammatory response [42]. In spinocerebellar ataxia (SCA) mouse models, the degree of microglia activation follows disease progression and correlates with the severity of damage and neuronal loss within the affected region of the CNS [43]. Microglia activation has been observed in several neurodegenerative diseases including Alzheimer's [44], Parkinson's [45], and amyotrophic lateral sclerosis [46, 47]; therefore, a potential role of microglia activation in equine CA was expected. In addition, microglia activation detection provides a molecular marker for severity of the CA phenotype, where its sensitivity and activation can appear before neuronal death [43].

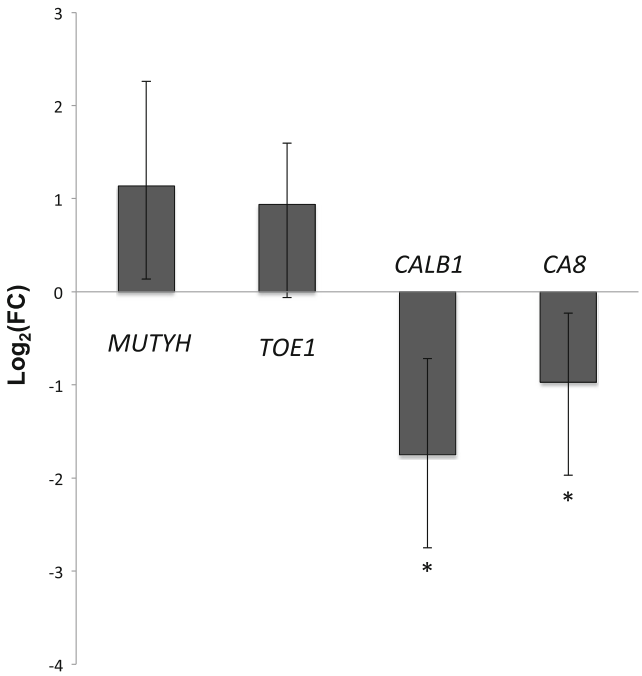

Fig. 6 qPCR results showing gene expression in CA-affected horses ($n = 4$) in terms of Log2(foldchange) relative to unaffected ($n = 5$) of *TOE1, MUTYH, CALB1*, and *CA8*. *CALB1* and *CA8* were significantly downregulated. *Error bars* are mean ± SEM, $^*P < 0.01$; Mann-Whitney *U* test for each gene

The second group of DE genes was down-regulated in CA-affected horses and pertained to calcium homeostasis. Panther overrepresentation tests also confirmed an enrichment in genes pertaining to calcium homeostasis in our 151 DE genes [34], including the genes *CA8*, *ITPR1*, and *TRPC3*. In humans, *CA8* has been affiliated with an autosomal recessive ataxia [48] with *ITPR1* as its allosteric inhibitor, mediating excitatory calcium signaling [49]. Purkinje neurons experience large influxes of calcium due to their excitatory rapid firing; therefore, genes pertaining to calcium regulation are often Purkinje neuron markers [50]. Subsequently, a loss of Purkinje neurons, as seen in CA, would cause a decrease in expression of genes associated with calcium homeostasis. Therefore, the down-regulation of calcium homeostasis-associated genes is most likely a consequence of Purkinje neuron loss. Along with the Purkinje-specific calcium homeostasis markers, genes involved in calcium homeostasis, but localizing to other cell types in the cerebellum, were also found to be differentially up-regulated. These genes include Vav Guanine Nucleotide Exchange Factor 1 (*VAV1*) and Rac/Cdc42 Guanine Nucleotide Exchange Factor 6 (*ARHGEF6*), which localize to granular neurons according to Allen Brain Atlas [36], and Pleckstrin Homology And RhoGEF Domain Containing G1 (*PLEKHG1*) and Calcium And Integrin Binding 1 (*CIB1*), which localize to the granular layer and appears to demonstrate some enrichment in the Bergmann glia according to Allen Brain Atlas [36] (Supplementary Table 9). These genes indicate an enrichment of gene expression, including calcium-related gene expression, from other cell types such as Bergmann glia and granular neurons with the absence of Purkinje neurons.

The results of unsupervised gene clustering were consistent with the DE analysis in that expression patterns of calcium homeostasis gene markers were affiliated with the CA-phenotype and clustered with the Purkinje neuron markers *CALB1* and Purkinje cell protein 4 (*PCP4*). However, the unsupervised

Table 2 The top significant Panther overrepresentation results regarding biological processes and molecular functions. The Gene Ontology (GO) ID and *p* value are in parenthesis

Input gene list	Biological processes	Molecular function
151 DE genes	- Cellular calcium ion homeostasis (GO:0006874, 3.28E-2)	- Nucleic acid binding (GO:0003676, 0.0253)
	- Homeostatic process (GO:0042592, 1.69E-2)	
Purkinje module	- DNA repair (GO:0006281, 8.15E-5)	- mRNA binding (GO:0003729, 1.09E-3)
	- mRNA processing (GO:0006397, 9.28E-8)	- RNA binding (GO:0003723, 4.67E-9)
Module 5	- Cellular protein metabolic process (GO:0044267, 1.63E-3)	- Cytochrome-c oxidase activity (GO:0004129, 4.28E-2)
	- Cellular metabolic process (GO:0044237, 1.64E-9)	- Ribonucleoside binding (GO:0032549, 4.10E-2).

method revealed this type of clustering with more extensive gene associations and suggests co-expression or co-regulation of these two gene sets, which can be extrapolated to calcium homeostasis gene expression being down-regulated and correlating to the decrease in Purkinje neuron population. The distinction of which neuronal population is expressing the down-regulation of this pathway becomes particularly important in cerebellar cortex slices, where there are other neuronal populations housed within the tissue, most notably, the more abundant granular neurons, which make up a larger proportion (granular layer) of the cerebellar cortex compared to the single layer of Purkinje neurons [51]. Despite not having the samples prepared in a manner conducive for microdissection of Purkinje cells, computational methods employing clustering and weighted gene networks can, to a certain degree, mitigate this deficit [30, 41]. Others have employed similar approaches pertaining to the cerebellum [30] and microglia populations throughout the brain [41], with the conclusion that such parsing of gene expression is a necessity when performing RNA-seq on a sample with heterogenous cell populations.

While unsupervised clustering established which gene expression patterns were correlated with the CA-phenotype, ROAST analysis confirmed the trends in expression patterns of these genes. Due to the loss of Purkinje neurons in CA-affected individuals, we observed significant down-regulation in expression of 40 % of Purkinje neuron gene markers without a corresponding down-regulation of granular neuron gene markers in CA-affected horses, confirming a primary loss of Purkinje neurons and their associated gene expression with a potential secondary loss of granular neurons in horses with CA.

Ataxias in other species involving failure to establish synapses with granular neurons has been seen in rabbits, where, as a result, there is apoptosis and irregular orientation of the Purkinje neurons [3]. In mice, there is also consequential degeneration of granular neuron axons and an increase in length of the remaining Purkinje neuron dendrites, with a reduction in dendritic spine branchlets [52]. This results in an incomplete loss of Purkinje neurons, with a small subset of intact Purkinje neurons receiving climbing or mossy fiber afferents. Our data does have a subset of Purkinje neuron markers that are up-regulated in CA-affected horses (Fig. 5, Supplementary Table 7), most interestingly APAF interacting protein (*APIP*) and inhibitor of DNA binding 2 (*ID2*). Both genes show strict expression in mouse Purkinje neurons in the cerebellum, according to Allen Brain Atlas [36], and have an increase in Log_2(foldchange) greater than 0.4 and FPKM values greater than 10 in both control and CA-affected horses; however, they do not have significant FDR values (Supplementary Table 7). *ID2* is recognized as associating with transcription factors that inhibit differentiation [53] and promotes proliferation of neuronal precursor cell types [54]. *APIP* inhibits two pathways leading to apoptosis: caspase-9-dependent apoptosis [55] and caspase-1-dependent pyroptosis [56]. Further and more targeted work confirming restricted expression of these two Purkinje neuron markers in equine Purkinje neurons and especially in the remaining Purkinje neurons of CA-affected horses is necessary. Also, a time course study evaluating time points before and after CA onset is necessary to deduce whether compromised synapses between granular and Purkinje neurons may be causing the ataxic phenotype and whether the Purkinje neurons exhibiting the *APIP* and *ID2* gene expression profiles confer some survival strategy for the Purkinje neurons remaining in CA-affected horses.

Secondarily, a large portion of the granular neuron markers were also significantly down-regulated (Supplementary Table 6), including Diacylglycerol Kinase Delta (*DGKD*), Solute Carrier Family 8 Member A2 (*SLC8A2*), and Sodium Voltage-Gated Channel Beta Subunit 2 (*SCN2B*). These markers are exclusively present in granular neurons according to the Allen Brain Atlas [36]. The down-regulation of these granular neuron gene markers suggests a potential associated loss of granular neurons in CA-affected horses. Further work focusing on the time course of neurodegeneration during CA would need to be done to confirm that granular neuron loss is secondary; however, due to husbandry, gestation, and opportunistic sample collection in horses, time course experiments are not feasible. Further work should be done in a mouse model integrating the proposed CA-SNP to better describe the molecular mechanisms driving CA disease progression in the horse.

Although differential expression of *MUTYH* and *TOE1* was not detected in the RNA-seq analyses, this study improved

annotation of their gene structure, demonstrated their potential involvement in cerebellar pathways, and, because of the similar expression levels of *TOE1* and *MUTYH* in horse cerebellar cortex, stimulated renewed consideration for *TOE1* as a candidate gene for CA. The limited timeframe and sample size prevent us from detecting developmental-specific differential expression of either of these candidate genes, but we were able to observe resulting expression patterns that may ensue from CA-SNP compromised *TOE1* (CA-*TOE1*). The cytoplasmic P-bodies associated with *TOE1* are known to respond to synaptic activity [15], and thus, further work characterizing the CA-*TOE1* interactions with other proteins and P-bodies as well as localization of *TOE1* in CA-affected horses must be done to explore the possibility of *TOE1* involvement in CA. Additionally, characterizing the subset of Purkinje neurons that remain after CA onset via immunohistochemistry and the genetic candidates *APIP* and *ID2* is required to confirm any selective subpopulation of Purkinje neurons that may not be affected in CA.

Conclusion

The early onset of equine CA suggests a problem in maturation of the Purkinje neuron circuitry during development that results in ataxia without a complete loss of Purkinje neurons. An analysis of molecular pathways preceding onset of CA is necessary to determine causal factors; however, we have characterized the loss Purkinje neurons in CA-affected horses by significant down-regulation of calcium-mediated signaling in the cerebellar cortex with concurrent up-regulation of microglia markers and a potential secondary loss of granular neuron markers.

Acknowledgments This project was funded in part by the Arabian Horse Foundation as well as the Animal Science Department and Veterinary Genetics Laboratory, University of California, Davis. Samples were provided by private donors and University of Pennsylavania, Colorado State University, and University of Kentucky. Additional acknowledgements to Dr. Tamer Mansour for his valuable advice concerning bioinformatics and Dr. Rebecca Bellone for overall critique of the project.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. de Lahunta A. Abiotrophy in domestic animals: a review. Can J Vet Res. 1990;54(1):65–76.
2. Koehler JW, Newcomer BW, Holland M, Caldwell JM. A novel inherited cerebellar abiotrophy in a cohort of related goats. J Comp Pathol. 2015;153(2–3):135–9.
3. Sato J, Yamada N, Kobayashi R, Tsuchitani M, Kobayashi Y. Morphometric analysis of progressive changes in hereditary cerebellar cortical degenerative disease (abiotrophy) in rabbits caused by abnormal synaptogenesis. J Toxicol Pathol. 2015;28(2):73–8.
4. Shearman JR, Cook RW, McCowan C, Fletcher JL, Taylor RM, Wilton AN. Mapping cerebellar abiotrophy in Australian kelpies. Anim Genet. 2011;42(6):675–8.
5. DeBowes RM, Leipold HW, Turner-Beatty M. Cerebellar abiotrophy. Vet Clin North Am Equine Pract. 1987;3(2):345–52.
6. Blanco A, Moyano R, Vivo J, Flores-Acuna R, Molina A, Blanco C, et al. Purkinje cell apoptosis in Arabian horses with cerebellar abiotrophy. J Vet Med A Physiol Pathol Clin Med. 2006;53(6):286–7.
7. Cavalleri JM, Metzger J, Hellige M, Lampe V, Stuckenschneider K, Tipold A, et al. Morphometric magnetic resonance imaging and genetic testing in cerebellar abiotrophy in Arabian horses. BMC Vet Res. 2013;9:105.
8. Palmer AC, Blakemore WF, Cook WR, Platt H, Whitwell KE. Cerebellar hypoplasia and degeneration in the young Arab horse: clinical and neuropathological features. Vet Rec. 1973;93(3):62–6.
9. Brault LS, Cooper CA, Famula TR, Murray JD, Penedo MC. Mapping of equine cerebellar abiotrophy to ECA2 and identification of a potential causative mutation affecting expression of MUTYH. Genomics. 2011;97(2):121–9.
10. Brault LS, Famula TR, Penedo MC. Inheritance of cerebellar abiotrophy in Arabians. Am J Vet Res. 2011;72(7):940–4.
11. Wagner E, Clement SL, Lykke-Andersen J. An unconventional human Ccr4-Caf1 deadenylase complex in nuclear cajal bodies. Mol Cell Biol. 2007;27(5):1686–95.
12. Zheng D, Ezzeddine N, Chen CY, Zhu W, He X, Shyu AB. Deadenylation is prerequisite for P-body formation and mRNA decay in mammalian cells. J Cell Biol. 2008;182(1):89–101.
13. Machyna M, Heyn P, Neugebauer KM. Cajal bodies: where form meets function. Wiley Interdiscip Rev RNA. 2013;4(1):17–34.
14. Parker R, Sheth UP. Bodies and the control of mRNA translation and degradation. Mol Cell. 2007;25(5):635–46.
15. Cougot N, Bhattacharyya SN, Tapia-Arancibia L, Bordonne R, Filipowicz W, Bertrand E, et al. Dendrites of mammalian neurons contain specialized P-body-like structures that respond to neuronal activation. J Neurosci. 2008;28(51):13793–804.
16. Baltanas FC, Casafont I, Weruaga E, Alonso JR, Berciano MT, Lafarga M. Nucleolar disruption and cajal body disassembly are nuclear hallmarks of DNA damage-induced neurodegeneration in purkinje cells. Brain Pathol. 2011;21(4):374–88.
17. Oka S, Ohno M, Tsuchimoto D, Sakumi K, Furuichi M, Nakabeppu Y. Two distinct pathways of cell death triggered by oxidative damage to nuclear and mitochondrial DNAs. EMBO J. 2008;27(2):421–32.
18. Lee HM, Hu Z, Ma H, Greeley Jr GH, Wang C, Englander EW. Developmental changes in expression and subcellular localization of the DNA repair glycosylase, MYH, in the rat brain. J Neurochem. 2004;88(2):394–400.
19. Sheng Z, Oka S, Tsuchimoto D, Abolhassani N, Nomaru H, Sakumi K, et al. 8-Oxoguanine causes neurodegeneration during MUTYH-mediated DNA base excision repair. J Clin Invest. 2012;122(12):4344–61.
20. Plotz G, Casper M, Raedle J, Hinrichsen I, Heckel V, Brieger A, et al. MUTYH gene expression and alternative splicing in controls and polyposis patients. Hum Mutat. 2012;33(7):1067–74.
21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
22. Wu D, Lim E, Vaillant F, Asselin-Labat ML, Visvader JE, Smyth GK. ROAST: rotation gene set tests for complex microarray experiments. Bioinformatics. 2010;26(17):2176–82.
23. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, et al. Analysis of the mouse transcriptome based on functional

annotation of 60,770 full-length cDNAs. Nature. 2002;420(6915):563–73.
24. Joshi NAJNF. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files 2011 [cited (Version 1.33)]. Software]. Available from: https://github.com/najoshi/sickle.
25. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14(4):R36.
26. Bray N, Pimentel, H., Melsted, P. & Lior, Pachter. Near-optimal RNA-Seq quantification. 2015;arXiv:1505.02710.
27. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013;31(1):46–53.
28. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
29. Team RDC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistcal Computing; 2010.
30. Kuhn A, Kumar A, Beilina A, Dillman A, Cookson MR, Singleton AB. Cell population-specific expression analysis of human cerebellum. BMC Genomics. 2012:13:610.
31. Kirsch L, Liscovitch N, Chechik G. Localizing genes to cerebellar layers by classifying ISH images. PLoS Comput Biol. 2012;8(12):e1002790.
32. Bettencourt C, Ryten M, Forabosco P, Schorge S, Hersheson J, Hardy J, et al. Insights from cerebellar transcriptomic analysis into the pathogenesis of ataxia. JAMA Neurol. 2014;71(7):831–9.
33. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Res. 2016;44(D1):D336–42.
34. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nat Protoc. 2013;8(8):1551–66.
35. D'Souza CA, Chopra V, Varhol R, Xie YY, Bohacec S, Zhao Y, et al. Identification of a set of genes showing regionally enriched expression in the mouse brain. BMC Neurosci. 2008;9:66.
36. Shen EH, Overly CC, Jones AR. The Allen human brain atlas: comprehensive gene expression mapping of the human brain. Trends Neurosci. 2012;35(12):711–4.
37. Biolatti C, Gianella P, Capucchio MT, Borrelli A, D'Angelo A. Late onset and rapid progression of cerebellar abiotrophy in a domestic shorthair cat. J Small Anim Pract. 2010;51(2):123–6.
38. Forman OP, De Risio L, Matiasek K, Platt S, Mellersh C. Spinocerebellar ataxia in the Italian Spinone dog is associated with an intronic GAA repeat expansion in ITPR1. Mamm Genome. 2015;26(1–2):108–17.
39. Sato J, Sasaki S, Yamada N, Tsuchitani M. Hereditary cerebellar degenerative disease (cerebellar cortical abiotrophy) in rabbits. Vet Pathol. 2012;49(4):621–8.
40. Whittington RJ, Morton AG, Kennedy DJ. Cerebellar abiotrophy in crossbred cattle. Aust Vet J. 1989;66(1):12–5.
41. Forabosco P, Ramasamy A, Trabzuni D, Walker R, Smith C, Bras J, et al. Insights into TREM2 biology by network analysis of human brain gene expression data. Neurobiol Aging. 2013;34(12):2699–714.
42. Block ML, Hong JS. Microglia and inflammation-mediated neurodegeneration: multiple triggers with a common mechanism. Prog Neurobiol. 2005;76(2):77–98.
43. Cvetanovic M, Ingram M, Orr H, Opal P. Early activation of microglia and astrocytes in mouse models of spinocerebellar ataxia type 1. Neuroscience. 2015;289:289–99.
44. Guillot-Sestier MV, Doty KR, Gate D, Rodriguez Jr J, Leung BP, Rezai-Zadeh K, et al. Il10 deficiency rebalances innate immunity to mitigate Alzheimer-like pathology. Neuron. 2015;85(3):534–48.
45. LE F, Tirolo C, Testa N, Caniglia S, Morale MC, Marchetti B. Glia as a turning point in the therapeutic strategy of Parkinson's disease. CNS Neurol Disord Drug Targets. 2010;9(3):349–72.
46. Sultan M, Amstislavskiy V, Risch T, Schuette M, Dokel S, Ralser M, et al. Influence of RNA extraction methods and library selection schemes on RNA-seq data. BMC Genomics. 2014;15:675.
47. Papadimitriou D, Le Verche V, Jacquier A, Ikiz B, Przedborski S, Re DB. Inflammation in ALS and SMA: sorting out the good from the evil. Neurobiol Dis. 2010;37(3):493–502.
48. Kaya N, Aldhalaan H, Al-Younes B, Colak D, Shuaib T, Al-Mohaileb F, et al. Phenotypical spectrum of cerebellar ataxia associated with a novel mutation in the CA8 gene, encoding carbonic anhydrase (CA) VIII. Am J Med Genet B Neuropsychiatr Genet. 2011;156B(7):826–34.
49. Hirota J, Ando H, Hamada K, Mikoshiba K. Carbonic anhydrase-related protein is a novel binding protein for inositol 1,4,5-trisphosphate receptor type 1. Biochem J. 2003;372(Pt 2):435–41.
50. Okubo Y, Suzuki J, Kanemaru K, Nakamura N, Shibata T, Iino M. Visualization of Ca2+ filling mechanisms upon synaptic inputs in the endoplasmic reticulum of cerebellar Purkinje cells. J Neurosci. 2015;35(48):15837–46.
51. Paxinos G. Cerebellum and Cerebellar Connections. In: Science E, editor. The Rat Nervous System. 4th ed. Burlington: Elsevier Science; 2014. p. 1053.
52. Anderson WA, Flumerfelt BA. Long-term effects of parallel fiber loss in the cerebellar cortex of the adult and weanling rat. Brain Res. 1986;383(1–2):245–61.
53. Neuman T, Keen A, Zuber MX, Kristjansson GI, Gruss P, Nornes HO. Neuronal expression of regulatory helix-loop-helix factor Id2 gene in mouse. Dev Biol. 1993;160(1):186–95.
54. Sullivan JM, Havrda MC, Kettenbach AN, Paolella BR, Zhang Z, Gerber SA, et al. Phosphorylation regulates Id2 degradation and mediates the proliferation of neural precursor cells. Stem Cells. 2016;34(5):1321–31.
55. Cho DH, Hong YM, Lee HJ, Woo HN, Pyo JO, Mak TW, et al. Induced inhibition of ischemic/hypoxic injury by APIP, a novel Apaf-1-interacting protein. J Biol Chem. 2004;279(38):39942–50.
56. Ko DC, Gamazon ER, Shukla KP, Pfuetzner RA, Whittington D, Holden TD, et al. Functional genetic screen of human diversity reveals that a methionine salvage enzyme regulates inflammatory cell death. Proc Natl Acad Sci U S A. 2012;109(35):E2343–52.