Journal of Genetic Mutation Disorders
Volume 1 | Issue 1

Research Article Open Access

Sequence Variant in the TRIM39-RPP21 Gene Readthrough is Shared Across a Cohort of Arabian Foals Diagnosed with Juvenile Idiopathic Epilepsy

Polani S[1], Dean M[2], Lichter-Peled A[1], Hendrickson S[3], Tsang S[4], Fang X[5], Feng Y[5], Qiao W[5], Avni G[6] and Kahila Bar-Gal G[*1]

[1]Koret School of Veterinary Medicine, The Robert H. Smith Faculty of Agriculture, Food and Environmental Sciences, The Hebrew University of Jerusalem, Rehovot, Israel
[2]National Cancer Institute, Division of Cancer Epidemiology & Genetics, Laboratory of Translational Genomics, USA
[3]Department of Biology, Shepherd University, Shepherdstown, USA
[4]BioMatrix LLC, USA
[5]BGI-Shenzhen, Shenzhen, China
[6]Medisoos Equine Clinic, Kibutz Magal, Israel

[*]**Corresponding author:** Kahila Bar-Gal G, Koret School of Veterinary Medicine, The Robert H. Smith faculty of Agriculture, Food and Environmental Sciences, The Hebrew University of Jerusalem, Rehovot, POB 12, Israel 76100, E-mail: gila.kahila@mail.huji.ac.il

Citation: Polani S, Dean M, Lichter-Peled A, Hendrickson S, Tsang S, et al. (2018) Sequence Variant in the TRIM39-RPP21 Gene Readthrough is Shared across a Cohort of Arabian Foals Diagnosed with Juvenile Idiopathic Epilepsy. J Gene Mutat 1(1): 103

Abstract

Juvenile idiopathic epilepsy (JIE) is a self-limiting neurological disorder with a suspected genetic predisposition affecting young Arabian foals of the Egyptian lineage. The condition is characterized by tonic-clonic seizures with intermittent post-ictal blindness, in which most incidents are sporadic and unrecognized. This study aimed to identify genetic components shared across a local cohort of Arabian foals diagnosed with JIE via a combined whole genome and targeted resequencing approach: Initial whole genome comparisons between a small cohort of nine diagnosed foals (cases) and 27 controls from other horse breeds identified variants uniquely shared amongst the case cohort. Further validation via targeted resequencing of these variants, that pertain to non-intergenic regions, on additional eleven case individuals revealed a single 19bp deletion coupled with a triple-C insertion (Δ19InsCCC) within the TRIM39-RPP21 gene readthrough that was uniquely shared across all case individuals, and absent from three additional Arabian controls. Furthermore, we have confirmed recent findings refuting potential linkage between JIE and other inherited diseases in the Arabian lineage, and refuted the potential linkage between JIE and genes predisposing a similar disorder in human newborns. This is the first study to report a genetic variant to be shared in a sub-population cohort of Arabian foals diagnosed with JIE. Further evaluation of the sensitivity and specificity of the Δ19InsCCC allele within additional cohorts of the Arabian horse is warranted in order to validate its credibility as a marker for JIE, and to ascertain whether it has been introduced into other horse breeds by Arabian ancestry.

Keywords: Arabian horse; juvenile idiopathic epilepsy; whole genome sequencing; targeted resequencing; breeding

Introduction

Juvenile idiopathic epilepsy (JIE) in Arabian foals is a self-limiting neurological disorder, manifested by recurrent tonic-clonic seizures with normal gestation and no known precipitating factors. Previous studies have reported high prevalence of JIE within the Egyptian lineage among the Arabian horse breed, suggesting that a genetic predisposition underlies the disease [1–3]. Although idiopathic epilepsy has been recognized in only a minute of seizure cases [2], it is suspected that more such cases either go unnoticed [4] or are deliberately withheld from being disclosed due to the high economic value of the Arabian horse [3]. Such circumstances have stressed the need to provide an efficient tool to identify potential carriers in order to promote cautious and responsible breeding thereby reducing the spread of the disease.

Few previous efforts were made in order to unravel the genetic component(s) predisposing JIE. Retrospective analysis of medical records pertaining to 22 cases postulated a dominant mode of inheritance [1], whereas a pedigree analysis of 12 case individuals and their associated relatives postulated a recessive mode of inheritance [3]. Furthermore, hypotheses of possible linkage to Lavender

Foal Syndrome (LFS), another genetic disease segregating within the Arabian breed, were raised [5,6], and recently refuted [3,7]. Attempts to address similar genetic predispositions underlying JIE and a phenotypically similar syndrome in humans, namely benign familial neonatal convulsions/syndrome (BFNC/BFNS), were also made, but with no conclusive remarks [3].

Due to these inconsistencies and pitfalls, a need for a more robust approach was warranted. Consequently, with the increasing success of genome wide association studies (GWAS) in resolving simple mendelian disorders in various horse breeds [6,8–10], a similar approach using the Equine SNP50 beadchip (Illumina®, San Diego, CA) was undertaken by the authors of this manuscript in order to identify potential genetic variant(s) associated with 12 case individuals among a cohort of 35 controls. However, no significant association could be identified [11].

Although undoubtedly successful, such GWA studies are commonly based on the common-disease common-variant model [12] by which a disease is associated with a single causative variant segregating with a minor allele frequency (MAF) of >5% [13]. In the case of epileptic syndromes with underlying genetic components, studies implementing such an approach may not prove successful due to the complex nature of these syndromes, which are commonly associated with either high locus heterogeneity and/or improper stratification of the cases and controls [13–15]. In the case of JIE, such a stratification is a grueling challenge. Proper diagnosis of JIE is primarily performed on clinical grounds and exclusion of other possible clinicopathological etiologies [1,2]. Although use of electroencapholgraphy (EEG) has been reported in selected cases as part of the diagnostic process [1,7] this procedure is usually impractical in the majority of cases due to the long inter-ictal periods and applied handling of the foals, essentially remaining a more common practice in human medicine. In respect to identifying true controls – the task is even more challenging. The notion that cases may go unnoticed [4] or may not be deliberately disclosed due to their economic value [3], increases the risk for dispersion of the disease within the population and reduces the chances of identifying true Arabian controls. Hence, for proper stratification, particular emphasis should be given to both medical and family history of control candidates.

The advances in next-generation sequencing (NGS) combined with contemporary bioinformatics approaches have become methods of choice in the study of complex genetic disorders [13,16]. Moreover, such methods have been particularly efficient in dissecting neurological diseases, especially when a certain variant may be segregating in a small number of relatives when large pedigrees are not always available or when a certain variant is segregating in small number of relatives [16,17] – circumstances that may be common in the equine industry. In this sense, whole genome sequencing (WGS) has indeed become a primary method for the study of epilepsy, as emphasized by the foundation of a number of highly-funded initiatives, e.g. the NIH's Epi4K and Epilepsy-Phenome-Genome projects, and the Welcome Trust's WGS500 project - all aimed to expand the genomic knowledge underlying epilepsy by sequencing genomes of thousands of epileptic human patients [14,18].

Owing to this, and considering the growing availability of equine genomes in the public domain [19–25], it is almost warranted that similar approaches be used in the study of JIE. Henceforth, this study aimed to identify genetic components associated with a local cohort of foals in Israel diagnosed with JIE via a combined whole genome and targeted resequencing approach. Initial comparative analyses between cases and control individuals identified a dataset comprising both SNVs and INDELs uniquely shared amongst a cohort of clinically diagnosed case individuals. Further validation of the above dataset on a larger set of case individuals revealed a single mutation uniquely shared across all case individuals.

Methods

Animal Welfare and Ethics Statement

This study utilized residual blood samples that were obtained for clinical use, and thus no additional discomfort to of the studied horses specifically for the purpose of this study were instigated. The use of the residual blood samples for the purpose of this study was approved by the Hebrew University Agricultural Faculty's Animal Care and Use Committee under permit number AG-23476-07.

Horses

A total of 22 subjects were sampled throughout the study in two cohorts: The first cohort ("screening cohort") was used for screening for variants uniquely shared across local Israeli foals diagnosed with JIE; The second cohort ("validation cohort") was used for validation of the variants identified in the first cohort across a larger set of case individuals (Table 1). Diagnosis of JIE was performed via attending veterinary physicians primarily on clinical grounds and exclusion of other possible clinicopathological etiologies as previously described [1–3].

The first screening cohort included a total of 10 individuals comprised of nine case (three males, six females; mean age: 2 months) and a single control (male) individual (Table 1). The control individual was carefully selected ensuring not to include any history of seizure-like manifests, blindness and/or events of unexplained facial trauma to suggest previous epileptic episodes until 12 months of age, and no history of seizure-like manifests within two generations back in his pedigree from both his sire's and dam's side. The second validation cohort included 12 additional case (2 males, 10 females; mean age: 2 months; Table 1). All studied individuals were purebred Arabian horses pertaining to the Egyptian lineage raised in Israel (Table 1). Although the identity of all sampled

horses was made available, those data are not provided in this manuscript to protect confidentiality. Furthermore, throughout the study, samples were coded to protect the anonymity of participating owners and farms.

(a) "Screening" cohort:

Sample ID	Status	Breed	Gender	Age (Months)
Arabian1	Case	Arabian	M	1
Arabian2	Case	Arabian	M	2
Arabian3	Case	Arabian	F	1
Arabian4	Case	Arabian	F	1
Arabian5	Case	Arabian	M	3
Arabian6	Case	Arabian	F	2
Arabian7	Case	Arabian	F	2
Arabian8	Case	Arabian	F	3.5
Arabian9	Case	Arabian	F	4
ArabianControl	Control	Arabian	M	Adult

(b)"Validation" cohort:

Sample ID	Status	Breed	Gender	Age
Arabian10	Case	Arabian	M	1.5 months
Arabian11	Case	Arabian	F	1 month
Arabian12	Case	Arabian	F	2 months
Arabian13	Case	Arabian	F	2 months
Arabian14	Case	Arabian	F	4 months
Arabian15	Case	Arabian	F	3 months
Arabian16	Case	Arabian	F	1.5 months
Arabian17	Case	Arabian	F	2 months
Arabian18	Case	Arabian	M	2 months
Arabain19	Case	Arabian	F	3 months
Arabian20	Case	Arabian	F	2 months
Arabian21	Case	Arabian	F	3 months

All Arabian horses are of the Egyptian lineage

Table 1: Sample data of the studied individuals

Sample Collection

Residual blood samples obtained for clinical use were studied. All samples were sterilely collected from the Jugular vein using an 18G needle and a 10ml syringe by the attending veterinarians. Blood was subsequently transferred into 3ml tubes containing 7.5% of EDTA anticoagulant. Tubes containing residual blood samples were coded and stored at 4 °C for DNA extraction.

DNA Extraction, Quantification and Validation

Blood samples were subjected to DNA extraction using a phenol-chloroform based method [26]. Quantifications of DNA extracts were made via the NanoDrop 2000 spectrophotometer (Thermo Scientific™). All extracts were diluted to a final concentration of 50ng/ul and stored at 4 °C for further processing.

Library Constructions and Whole Genome Sequencing

DNA samples were sheared randomly via a Covaris® AFA™-based platform to obtain a fragment length of 500bp following gel purification. Subsequently, purified products were subjected to end-repair, 3'-A-tailing and adapter ligation. Following product size selection via an additional gel purification step, a PCR-enrichment step was performed and objective length library fragments were purified and marked accordingly. All steps were performed according to the manufacturer's guidelines (Illumina®). Library quality control was performed by Agilent 2100 Bioanalyzer and ABI Step One Plus real-time PCR system. Finally, libraries passing quality control (QC) were subjected to paired-end (PE) sequencing via the Illumina Hiseq2000 platform in two consecutive rounds: (a) high-coverage WGS of two cases (Arabian1 and Arabian2) and the single control (Arabian Control) individual, for which 90bp PE reads were generated; (b) low-coverage WGS of the remaining 7 case individuals (Arabian3, Arabian4, Arabian5, Arabian6, Arabian7, Arabian8, Arabian9) for which 100bp PE reads were generated (Table S1).

Quality Control of Sequenced Reads

Quality assessment of the sequenced reads was performed using FastQC [27], and reads were subsequently subjected to the following filtering steps using in-house algorithms: (a) Removal of adapter sequences (b) Reads with QUAL scores <35 over 50% of their total length, or reads having >10% N's along the entire read length were dropped. Additionally, for the high-coverage WGS genomes, reads were also cut short in order to remove 2bp from the 5'-end and 3bp from the 3'-end.

Sequence Mapping and Genomic Assemblies

Sequence reads passing quality control filters were mapped to the EquCab2.67 reference assembly, excluding the mitochondrial reference sequence [28]. Indexing of the reference assembly was sought using SAMtools v0.1.19 FAIDX algorithm [29]. Sequence reads pertaining to each individual were aligned to the reference assembly using the Burrows-Wheeler Aligner v0.5.9 [30] with the following adjustments: (a) maximum number of gap extensions set to 50bp (b) no Indel shall be placed within 15bp from the end of a sequence (c) QUAL threshold for trimming down reads to 35bp set to 10. Duplicate reads within the resulting SAM files were marked and removed using Picard Tools' v1.82 MarkDuplicates algorithm and were subsequently merged into a single SAM file per individual using MergeSamFiles algorithm (http://picard.sourceforge.net/). Resulting SAM files were converted to BAM file formats, sorted according to chromosomal positions, indexed and inspected for basic statistics accordingly using SAMtools' VIEW, SORT, INDEX and IDXSTATS algorithms [29]. Subsequently, genomic alignments were subjected to the following refinement steps in order to produce high quality BAM files: (a) local realignments were performed using Genome Analysis ToolKit's (GATK; GATK v2.2-5; [31,32]) RealignerTargetCreator and IndelRealigner algorithms; (b) Paired-end information was fixed accordingly using GATK's FixMateInformation algorithm. Quality assessments of genomic coverages and standard deviations of the final high quality BAM files were performed using Qualimap v0.7.1 [33]. Breadth of coverage (percent of assembly length per reference length) was calculated using SAMtools' depth algorithm [29].

Additional Control Genomes

Genomic assemblies of previously published modern equine purebreds available at the time of the study (early 2017) were retrieved from online resources in order to serve as additional controls. These include: (a) Quarter horse [20]; http://server1.intrepidbio.com/FeatureBrowser/ngsdatasetrecord/record?ngsrecord=6197673305) (b) Icelandic; (c) Norwegian Fjord; (d) Standardbred; (e) Newly sequenced version of Twilight, i.e. Thoroughbred [24]; http://geogenetics.ku.dk/publications/middle-pleistocene-omics). In cases where genomic assemblies were not available, raw sequence data was retrieved from the NCBI's BioProject database for: (a) Duelmener, (b) Sorraia, (c) two Hanoverian males [23]; collectively under accession: PRJNA230019; (d) Marwari [21]; accession: PRJNA246445; (e) Mongolian male and female [19]; accession: PRJNA233529; (f) Korean Jeju male and female ponies [22]; accession: PRJNA169102. For raw sequence data, genomic assemblies were carried out faithfully following the authors' detailed description within the relating publication. In order to further expand the control cohort, an additional Andean horse was included (Hendrickson, personal communication), and 12 additional genomes of purebred Franches-Montagnes were kindly donated on behalf of Prof. Orlando from the Centre of Geogenetics (Natural History Museum of Denmark, personal communication), amounting to a total of 27 non-Arabian controls pertaining to 13 horse breeds.

Although other Arabian horse genomes were also publically accessible, they were not utilized for the purpose of this study since their clinical history was not available and thus it was not feasible to classify them as carriers or controls.

Identification of copy-number variations (CNVs)

Control-FREEC [34] was used in order to identify CNV regions across all genomes of individuals within the screening cohort. In order to reduce false positive identification of regions associated with low mappability within the reference genome, custom tracks of such regions were generated manually using the GEM mappability tool [35], and implemented into the Control-FREEC pipeline. Parameters used included: (a) ploidy=2; (b) coefficientOfVariation=0.05; and (c) breakPointThreshold = 0.08. Regions <100bp were excluded, and CNVs uniquely shared across all case individuals and exempt from the Arabian control were identified using custom in-house algorithms. Verification of uniquely shared CNV-regions was sought manually by eye via GenomeBrowse® (Golden Helix) by comparing the relative pileup pattern of the regions across all cases and control individuals within the screening cohort. Uniquely-shared CNVs were considered true only if the relative pileup pattern within the designated region which harbored them differed significantly between the case cohort and all the control individuals. Identification the associated genes was also sought via GenomeBrowse®.

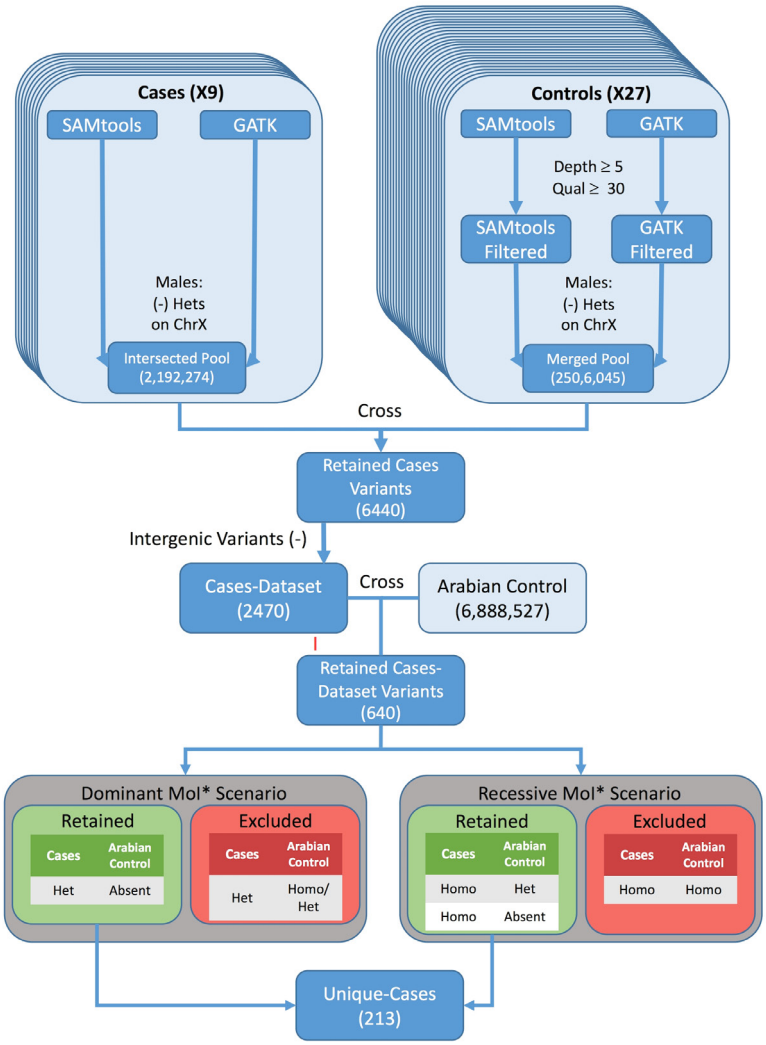
Variant Calling, Filtering and Crossing Strategy

Given this study aimed to identify casual variant(s) of a suspected genetic disease, a sensitive approach for variant identification was embraced, i.e. false positive calls were considered tolerable, whereas false negative calls were considered intolerable. Hence, two algorithms were used conjointly for variant calling, namely GATK's UnifiedGenotyper (v2.2-5; [31,32]) and SAMtools' mpileup (v0.1.18; [29]). In order to maximize the number of variant calls pertaining to the case individuals, no filtering parameters/steps were used during this phase, postulating that variants of low quality and irrelevancy to the study will be excluded later on in the pipeline when crossing the datasets of all case individuals. Nevertheless, in order to decrease potential false negatives which may

result from potential crossings with low-quality variants pertaining to the control individuals, variants with a depth <5 and quality scores <30 were excluded from the controls using GATK's SelectVariants algorithm.

Next, the two call sets per each individual were merged into a single dataset using GATK's CombineVariants algorithm. In cases where a variant call was shared by both sets, QUAL scores were prioritized to those apparent via GATK's UnifiedGenotyper. For male individuals, heterozygous calls pertaining to the X chromosome were excluded prior to final merges of the two datasets using GATK's SelectVariants.

Next, variant calls pertaining to each of the case individuals were merged into a single dataset using GATK's CombineVariants. Only variants shared across all case individuals were retained for further analysis using GATK's SelectVariants. Finally, this dataset was crossed with the datasets of all 27 non-Arabian controls via GATK's CombineVariants, and only variants uniquely shared across all nine case individuals (and were absent from all controls) were retained for further analysis using GATK's SelectVariants (Figure 1).



*MOI = Mode of Inheritance; ChrX = Chromosome X; Het = Heterozygote; Homo = Homozygote
Numbers in brakets represent number of retained variants.
Figure 1: Variant Calling and Filtering Strategy

Functional Genomics & Variant Annotations

Basic annotations and effect predictions of the variants pertaining the resulting dataset were conducted using snpEff v3.5d [36]. Individual VCF files were further filtered so to exclude variants relating to intergenic regions, using GNU-AWK based in-house algorithms (http://www.gnu.org/), essentially maintaining variants referring to exons, introns, 5'-UTR, 3'-UTR, upstream, and downstream regions. Synonymous variants were excluded from further analyses. Final gene identifiers (HGNC symbols) relating to each variant were obtained by cross-matching corresponding Ensembl gene IDs with human orthologous via Ensembl BioMart [37]. Henceforth the resulting dataset is designated 'Cases-Dataset'.

Further Filtration and Verification Strategies

In order to reduce the number of resulting variants, the Cases-Dataset was further crossed with the variant database of the Arabian control (Figure 1). However, owing to the elusive nature (in which breeders tend to withhold information of true cases) and undetermined mode of inheritance of the disease, it was hard to refute the slight possibility that the Arabian Control may still be a potential carrier, given a recessive mode of inheritance. Hence, five independent considerations were made when crossing the Cases-Dataset variants with those pertaining to the Arabian control: In case of a recessive mode of inheritance: (a) Homozygous variants in Cases-Dataset with homozygous counterparts in the Arabian control were excluded; (b) Homozygous variants in Cases-Dataset with heterozygous counterparts were retained; (c) Homozygous variants in Cases-Dataset with no counterparts in the Arabian control were retained; In case of a dominant mode of inheritance: (d) Heterozygous variants in Cases-Dataset with homozygous/heterozygous counterparts in the Arabian control were excluded (e) Heterozygous variants in Cases-Dataset with no counterparts in the Arabian control were retained. All cross inspections between the Cases-Dataset and the Arabian control were made manually by eye via GenomeBrowse® (Golden Helix) in order to verify that all retained variants meet the above criteria and that none of them were evident via the non-Arabian controls. Henceforth, the final resulting dataset is designated 'Unique-Cases' (Figure 1).

Validation of Candidate Variants Via Ion Torrent & Sanger Sequencing

A semi-conductor based targeted resequencing approach was used in order to validate the presence of the 213 variants within the Unique-Cases dataset across 12 additional cases individuals, i.e. the "validation cohort" (Table 1b). For this purpose, the Ion AmpliSeq™ Designer tool v5.2 was utilized for designing ultrahigh-multiplex primers. In order to maximize the specificity of the primer design, the designer algorithm needs to search across the largest available reference dataset. Hence, the complete EquCab2.67 reference genome was used and uploaded manually to the designer cloud. Next, a tapering design approach was used in order to identify the largest cohort of primer sets able to simultaneously amplify and sequence the maximum number of variants in a single pool, starting with a design based on the maximum target size (125bp) and for every following design the target sized was reduced by 10-15bp. A final cohort of 183 amplicons covering 97.6% of the variants in the Unique-Cases dataset (208/213 variants) was generated for further amplification and sequencing (Table S2). DNA of samples pertaining to the validation cohort were then subjected to library preparation, amplification and semi-conductor sequencing on an Ion Torrent PGM Sequencer following the manufacturer's guidelines (Thermo Fisher Scientific). Raw sequence reads were mapped to the reference genome using BWA as previously described. Mean depths of coverage per each target region of the resulting BAM files were obtained via SAMtools' bedcov [29], BEDtools' coverageBed [38] and in-house algorithms. Final inspections for the presence of the variants within the Unique-Cases dataset across the resulting BAM files were performed manually by eye via GenomeBrowse®.

For amplicons that failed to amplify, and regions for which the AmpliSeq™ Designer tool failed to identify target primers (Table S2), targeted PCR amplifications followed by Sanger sequencing was used in order to validate for the presence of the corresponding variants across the larger cohort of case individuals. Custom primers were designed using Primer3Plus and optimizing for Tm of 60 °C ([39]; Table S3). DNA amplifications, concurrent sequencing and manual inspections of corresponding chromatograms were performed using previously published protocols [40].

Results

Whole Genome Sequencing and Mapping Data

Genomic DNA was successfully sequenced using the Illumina HiSeq 2000 sequencer to generate a total of 6,725,185,309 DNA reads from all nine cases and a single control individual that followed this study's guidelines (see Methods section). Following all refinements steps, mean genomic coverages were 46.44x for the first round of sequenced genomes (Arabian1, Arabian2 & ArabianControl) and 13.48x for the second round (Arabian3, Arabian4, Arabian5, Arabian6, Arabian7, Arabian8, Arabian9), with an average breadth of coverage of 97.7% and a total average mapping rate of 96.83% for all 10 genomes. All sequence and genomic data are available via Table S1.

Identification of CNVs

CNV containing regions per each individual assembly were identified using Control-FREEC [34], and further crossed in order to identify shared CNV regions. A total of 147 regions were identified as uniquely shared across all case individuals and exempt from the Arabian control. Twenty-five of these regions were embedded within genes of which eight were of known function (*SWT1, SLC2A13, KCNQ1, CARS, ACSM5, COG6, HLA-DQB1, ADAM18;* Table S4). However, manual verification by eye via GenomeBrowse® (Golden Helix®; http://goldenhelix.com/products/GenomeBrowse/) failed to validate the presence of the shared pileup copy-number nature of these regions, essentially refuting their credibility.

Variant Call Data

Basic statistics including distribution of homozygous and heterozygous calls were obtained using vcftools v.0.1.11 [41], GNU-AKW, GNU-GREP and GNU-SED -based one-liner scripts (http://www.gnu.org/). All variant statistics are available in Table

S5. Both variant rates (average total, SNV and Indel rates: 0.00277, 0.00244 and 0.000325 respectively), and heterozygosity rates (average: 0.00146) of all sequenced Arabian genomes were similar to those apparent in other previously published horse genomes (Table S6; [21–23]). No trend differences in normalized total number of variants per chromosome lengths were evident between the cases and control individuals (Figure S1), refuting the possible association of genomic hotpots with the case-cohort.

Following implementation of variant intersections, annotations and filtration strategies, a final dataset of 213 variants was comprised for the Unique-Cases dataset, essentially representing variants uniquely shared across all case individuals (and absent from all controls) that do not pertain to intergenic regions (Figure 1, Figure S2, Table S7).

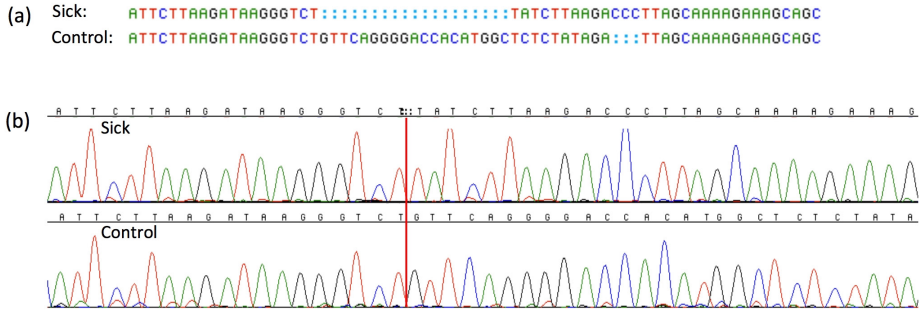
Functional Genomics

Functional predictions and distributions of the 213 variants uniquely shared across all cases within non-intergenic regions were inspected (Table S7 and Figure S2). A total of 71 equine genes (54 known protein coding, three uncharacterized protein coding genes, one miscRNA, eight pseudogenes, five processed pseudogenes) pertaining to 62 human orthologs were identified across the 213 variants, with *COL27A1, CUL5*, *FLT3* and *GBE1* as the top ranking genes to accumulate the highest number of variants (72, 31, 11 and nine variants, respectively) (Table S8).

Additional inspections for the prevalence of variants of known traits and diseases did not identify any disease-associated mutations shared across all individuals of the screening cohort, including LFS (chr1: 138235715C>Del). However, they were found to be unanimous for the chestnut coat color variant (chr3:36259552C>T in the *MC1R* gene) and the body size variants (chr6:81481065C>T in *HMGA2*; chr9:75550059T>C in ZFAT) (Table S9).

Validation Via Ion Torrent

A total of 170/183 (92.9%) amplicons corresponding to 182/213 (85.4%) of the total variants within the Unique-Cases dataset were successfully amplified across 11/12 individuals of the validation cohort via the Ion Torrent system (Tables 1b [Methods section] and S2). Mean read length was 200bp and the mean coverage per each region was 114.2 (Table S2). Manual inspections of the corresponding amplified regions via GenomeBrowse® (Golden Helix®; http://goldenhelix.com/products/GenomeBrowse/) identified a single 19bp deletion coupled with five ambiguities and a triple-C insertion (Δ19InsCCC; chr20:29542397-29542425: GTTCAGGGGACCACATGGCTCTCTATAGA>TATCTTAAGACCC; Gene Name: *TRIM39-RPP21*; Figure 2) that was uniquely shared across all successfully amplified case individuals (20 individuals) and exempt from all the control individuals including the Arabian control. Eleven case individuals were homozygous for the deletion (Arabian3, Arabian4, Arabian7, Arabian9, Arabian10, Arabian11, Arabian12, Arabian13, Arabian14, Arabian15, Arabian20), and nine were heterozygous (Arabian1, Arabian2, Arabian5, Arabian6, Arabian8, Arabian16, Arabian17, Arabian18, Arabian19). *No other variants that followed this study's filtration strategy for both dominant or recessive modes of inheritance and were simultaneously shared across all the larger cohort of case individuals were identified.*



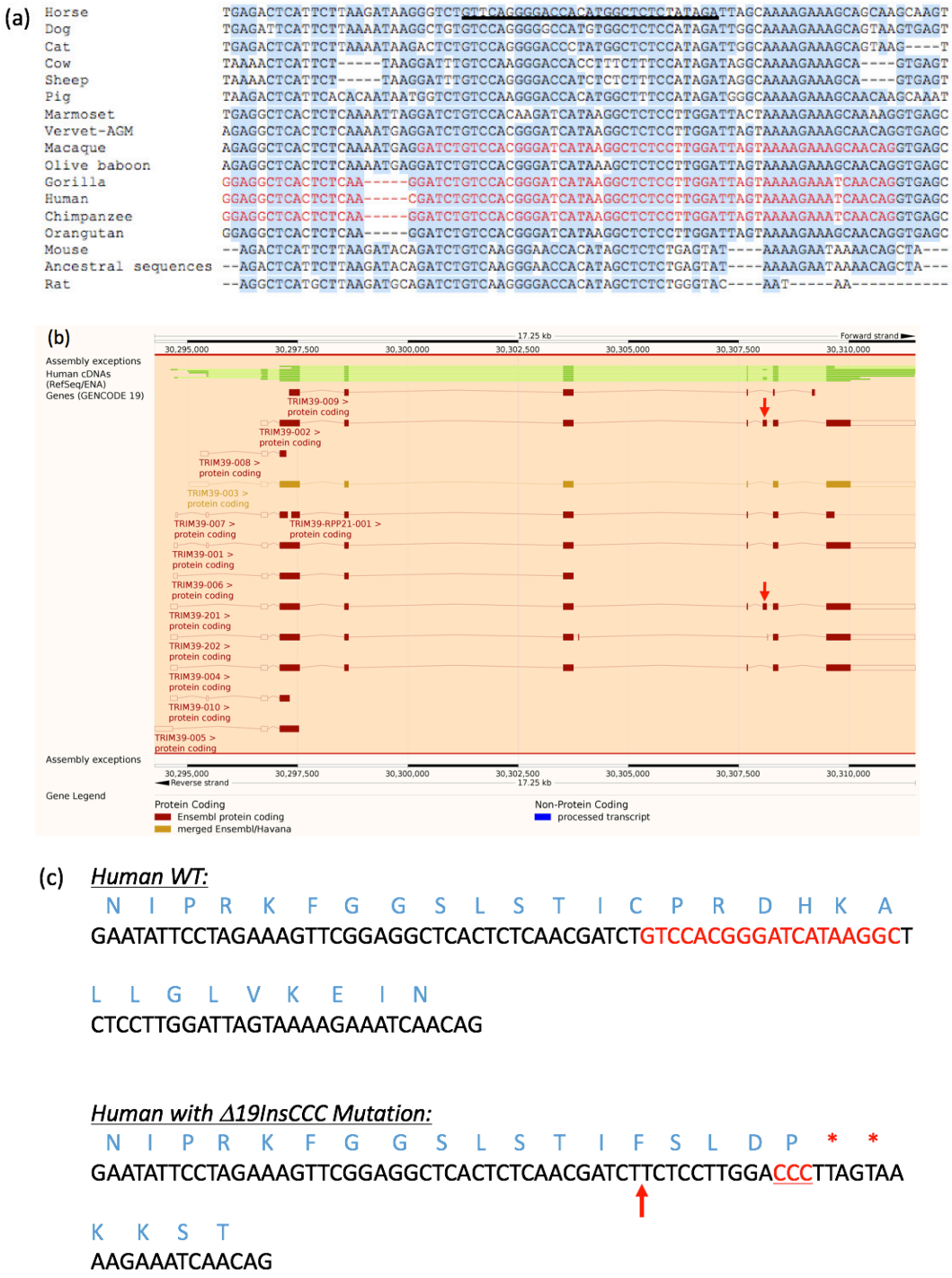
(a) Sequence alignment of the Δ19InsCCC mutation between a sick and a control individual.
(b) Aligned chromatograms. The 5'-end depicting the beginning of the mutation is marked with a red line.
Figure 2: Validation of The Δ19InsCCC mutation via Sanger sequencing

Of the remaining 31 variants that either failed to amplify via the Ion Torrent system or were not included in the sequencing process due to failure of the AmpliSeq Designer tool to identify target primers, 23 variants were successfully amplified by custom primers specifically designed via Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi; Table S3). None of these variants were shared across all the case individuals in the larger validation cohort, essentially refuting their candidacy as potential causative variants. For the remaining eight variants, we were unable to design custom primers due to (a) low mappability, i.e. poor assembly, of the target region in the reference genome; (b) lack of specificity of the resulting primers; (c) long N-stretches adjacent to the target region, thus essentially leaving these regions uninspected (Table S3).

Comparative Genomics

Comparative genomic analysis of the Δ19InsCCC mutation performed via Ensembl's comparative genomic resource [42] showed that the locus is highly conserved across other species of the Amniota vertebrates (Figure 3a) and was absent from their corresponding dbSNP databases, including that of the horse. Furthermore, although the Δ19insCCC corresponds to intron 4 of the *TRIM39-RPP21* gene in the horse, it was also found to correspond to exons 6 and 7 of two consecutive transcripts of the

TRIM39 gene in human (GRCh37 genomic region: chr6:30308090-30308108; Ensembl transcript IDs ENST00000396547.1 and ENST00000376656.4, respectively; Fig 3b,c). Further evaluation of the functional effect that the Δ19InsCCC mutation may have on the human orthologue was performed via PROVEAN v1.1.3 [43]. Results designated the mutation as a "frameshift" leading to a truncated protein due to the formation of premature termination within the resulting transcript (Figure 3c). Additional screenings throughout the horse genome via Ensembl and UCSC genome browsers failed to identify the presence of the separate *TRIM39* and *RPP21* genes.



(a) Multiple alignment of the Δ19InsCCC mutation region across 16 different species of the Amniota and the horse. Blue shadings represent highly conserved regions. Red sequences represent core exons regions within the designated species' genome.
(b) Regional overview of the TRIM39 gene in the human genome. Exons in which the Δ19InsCCC mutation resides are marked with red arrows (exons 6 and 7 of the human Ensembl transcripts ENST00000396547.1 and ENST00000376656.4, respectively).
(c) Genomic and translated protein sequences corresponding to the above exon sequences. Translated sequence is given in blue fonts above the first codon of each triplet. The orthologous 19bp deletion is marked in red fonts. The resulting truncated sequence is presented at the bottom, with the red arrow marking the beginning of the mutation and the triple-C insertion underlined in red. Red asterisk correspond to stop codons.
Figure 3: Comparative genomics of the Δ19InsCCC mutation gene region

Inspection of Public Equine RNA-Seq Data for Potential Transcripts With Exons Harboring the Δ19InsCCC Variant

Screenings were performed by BLAT searches of a query spanning the entire Δ19InsCCC region along with additional 30 bp of both the up- and downstream regions (chr20:29542367-29542455) using UCSC's Genome Browser's BLAT search tool (http://genome.ucsc.edu/index.html) across track hubs pertaining to the newly published "refined horse transcriptome" [44] and two additional assemblies [45,46]. Track hubs were adopted from the horse trans repository on GitHub (https://github.com/dib-lab/horse_trans) as described by Mansour and colleagues (2017; [44]). Results revealed no evidence of predicted exon-transcripts harboring the Δ19InsCCC variant or its reference counterpart (Figure S3). Furthermore, additional BLAT searches of the above query were also performed in a similar manner across two additional publically available transcriptomes of Arabian horse origin [47,48] representing 14 horses in total. Results revealed no evidence of reads harboring the Δ19InsCCC variant.

Sequence Verification of the Δ19InsCCC Mutation Via PCR and Validation of Additional Potential Controls

Primers targeting the Δ19InsCCC mutation were designed via Primer3Plus as in a similar manner to that mentioned herein [39]; Forward Primer: 5'-GATGAACTATAACAGTATCCCCAGA-3'; Reverse Primer: 5'-ATACTGAAGTCACCTCCATGGTCT-3'). PCR amplifications and concurrent sequencing were performed according to previously described protocols [40]. Results confirmed the presence of the Δ19InsCCC mutation among case individuals (Figure 2) and absence from the Arabian Control. Although DNA samples of additional "control" candidates were made available to us, we were either not able to validate their clinical history, and/or clinical data pertaining to their sire/dam throughout their pedigree, at least two generations back (see requirements in Methods section), essentially excluding their candidacy. However, we were able to identify two individuals that met the above criteria, in which trustworthy clinical data was made available only for their parents (sire and dam). PCRs and consequent sanger-sequencing validated the absence of the Δ19InsCCC variant from these two samples. Therefore, all three Arabian controls were absent for the Δ19InsCCC variant.

Discussion

This study is the first to apply parallel WGS and targeted resequencing in order to identify genetic markers associated with an inherited disorder in the horse. We have described a 19bp deletion coupled with a triple-C insertion (Δ19InsCCC) in *TRIM39-RPP21* gene readthrough that was found to be uniquely shared in a local cohort of Egyptian Arabian foals clinically diagnosed with JIE, and absent from a total of three "true" Arabian controls. Moreover, no evidence for additional genetic mutations associated with other inherited disorders in the Arabian lineage were found to be shared across the studied cohort, essentially refuting previous speculations concerning potential linkage between JIE and other inherited diseases within the breed, and thus confirming similar recent findings [5-7]. Furthermore, no mutations within genes associated with a similar inherited disorder in human newborns (i.e. Benign Familial Neonatal Convulsions associated with mutations in the *KCNQ2* and *KCNQ3* genes) were identified in the final Unique-Cases dataset, essentially refuting the potential linkage between JIE and the genes predisposing BFNC/BFNS [3].

This study was particularly challenging due to a number of issues: (a) The fact that all case individuals were of the same breed resulted in a high number of shared variants that warranted robust filtration strategies; (b) Given the elusive nature of the disease phenotype, identifying true Arabian controls with no previous history of seizure-like manifests, blindness and/or events of unexplained facial trauma to suggest previous epileptic episodes during its first year of life (and similarly throughout its family history) was increasingly complex, and resulted in only a single Arabian control that met the full requirements set-forth in this study (see Methods section); (c) The inconsistency in previous reports pertaining to the mode of inheritance of the disease [1,3] obscured filtering steps within the analysis pipeline. We have attempted to overcome these hurdles by: (a) Using a high number of non-Arabian controls in combination with the single Arabian control throughout the initial screening phase; (b) Refraining from using "presumed" Arabian control samples with no clear medical and family history (including those available in the public domain) in our initial screening phase in order to maintain proper stratification; (c) Adopting a unique and cautious variant filtering strategy aimed to retain variants that follow both a dominant and recessive modes of inheritance; (d) Validating the variants identified as uniquely-shared among the screening cohort on additional individuals, and (e) Validating the absence of the Δ19InsCCC variant across two additional Arabian Controls that partially met the requirements of this study, mounting to a total of three Arabian Controls used throughout this study.

Although the Δ19InsCCC variant was found within an intron of the equine *TRIM39-RPP21* readthrough, comparative genomic analysis identified it to be conserved across additional species of the Amniota and to correspond to a frameshift mutation in core exons of the orthologous TRIM39 gene of primates, including the human. This disparity, topped by the absence of both *TRIM39* and *RPP21* gene annotations in the horse genome and the lack of available transcript data to support evidence of exon-transcripts harboring the Δ19InsCCC variant or its reference counterpart, accentuates the different degree of curations and annotations that highly revised genomes, such as the human, may have relative to other poorly studied genomes. This notion is also emphasized by the high complexity portrayed by chromosomes 12, 20 and the unmapped regions implied by the high number of normalized variants identified in this study, topped by the high number of CNVs identified in previous studies [49,50]. Such phenomena emphasize not only the need for caution when investigating non-human genomes, but also warrant further enhancement of their genomic, annotative and curative levels. In this respect, efforts are being made to release an enhanced version of the equine

genome, EquCab3.0 [51], which will hopefully aid in bridging this gap.

Tripartite motif (TRIM) proteins are E3 ubiquitin ligases which are required as part of the ubiquitination pathway to target specific substrates for degradation by the 26S proteasome, a key regulatory process of many enzymes [53]. TRIMs have been found to regulate steady state levels of a number of ribonucleases, including Ribonuclease P Protein Subunit P21 (*RPP21*), which take part in modulating G1/S transitions in the cell-cycle and modulation of DNA repair associated with genotoxic stress [52,54,55]. Interestingly, they have been shown to be involved in innate immune responses [56], and especially in antiviral defense [57].

Although the pathogenic significance of the vast number of TRIMs is not fully determined, some members have been reported to be highly expressed throughout the nervous system and associated with epilepsy. Examples include TRIM3, TRIM8 and TRIM32 which have been tied to epilepsy via hippocampal plasticity, regulation of synaptic neurotransmitters, and immunomodulation, respectively [58-64]. Other TRIM proteins associated with neuromodulation include: *TRIM2* [65,66], *TRIM9* [67,68], *TRIM11* [69], *TRIM47* and *TRIM65* [70].

Pathological disorders associated with *TRIM39* are mainly associated with immune-mediated conditions, i.e. Behçet's syndrome [71], cutaneous lupus erythematosus [72] and inflammatory bowel disease [73]. In this respect, TRIM39 was recently found to negatively regulate the NFκB-mediated signaling pathway [74], well known for its immunomodulatory role [75]. It is also a crucial transcription factor for glial and neuronal cell function [76] and plays a key role in neurogenesis in both embryonic and adult brains [77]. Interestingly, previous studies have proposed contradicting mechanisms for the role of NFκB activity in modulating epileptogenesis. For example, Lubin and colleagues (2007) demonstrated that inhibition of NFκB significantly decreased seizure threshold in kainic acid (KA)-treated rats, suggesting a neuroprotective role for NFκB during epileptogenesis. On the other hand, Di and colleagues (2011) demonstrated that NFκB inhibition decreased seizure susceptibility in KA-induced seizure rats by attenuating the overexpression of P-glycoprotein, a key transport protein in the blood brain barrier [80].

Owing to *TRIM39*'s immunomodulatory role and given the prevalent representation of TRIMs within the nervous system, it is reasonable to hypothesize that the former may also have a key role in modulating brain enzymatic activity. However, the exact mode of action by which this complex phenomenon is achieved, its contribution to epileptogenesis in general and in Arabian foals in particular, is yet to be determined.

Although seemingly encouraging, it is important to note the limitations of this study. While the sample size of the case individuals used in this study is comparative to other researches of heritable disorders in the horse [1,81-83], the number of control individuals does not include representatives of all horse breeds. This instance resulted in a high number of variants uniquely shared across the case screening cohort of individuals following exclusion of the variants shared across the 27 non-Arabian controls and the single Arabian control. Although the downstream filtration strategy utilized in this study reduced the number of variants considerably, the course of the study was to focus on variants within the vicinity of annotated genes. Hence, we cannot refute the possible omission of other intergenic variants that may well be associated with JIE. However, one cannot ignore the notion that in order to thoroughly investigate of the possible association of intergenic variants to any inherited disorder, the number of case and control genomes need to be significantly increased, with both groups being of high heterogeneity. Failure to follow this strategy will ultimately result in variants relevant only to the studied cohort and not necessarily applicable to other sub-populations diagnosed with the same inherited disorder. For example, BFNC/BFNS in human newborns was originally thought to be predisposed exclusively by mutations in the potassium channel encoding *KCNQ2* gene on chromosome 20 [84-87]. Subsequently mutations in the *KCNQ3* gene on chromosome 9 [87,88] and an inversion on chromosome 5 [89], were identified as additional loci. Consequently, the syndrome was sub-classified according to the associated genes, i.e. BFNS1, 2 and 3. Given this study's designed approach, it is important to note that the Δ19InsCCC allele may very well be considered as a marker (and not necessarily a causative variant) for JIE relevant only to the studied population, and, as in the case of BFNC/BFNS, additional markers may be identified in other sub-populations of the Arabian horse. Hence, caution should be taken when considering implementation of the test described in this manuscript for diagnostic purposes.

Conclusions

The modern Arabian horse has endured a strong bottleneck effect due to the numerous backcrosses typical of its historical and modern-day breeding programs. This phenomenon has led to the preservation of unwanted genetic disorders [90-92], with JIE being one of six such disorders segregating within the Arabian breed [1,4,6,81,82,93,94]. This is the first study to identify genetic variants uniquely shared across a sub-population of local Arabian foals diagnosed with JIE. Further evaluation of the sensitivity and specificity of the Δ19InsCCC allele within the Arabian horse breed is warranted in order to validate its credibility as a marker for JIE in both local and other sub-populations, and to ascertain whether it has been introduced into other horse breeds by Arabian ancestry.

Funding

This work was supported by The Israel Science Foundation (ISF), grant #1365/10.

Acknowledgments

We would kindly like to thank Prof. Ludovic Orlando and Dr. Pablo Librado for their critical reading and kind donation of the

Franches-Montagnes horse genomes. We would also like to thank all the attending veterinary personnel at The Large Animal Department of The Hebrew University's Veterinary Teaching Hospital and all the referral veterinarians for their sincere dedication to the promotion of this study. We thank Dr. Amir Steinman for his clinical support of this project. We thank Prof. Michael J. Gutnick for his critical thinking and scientific recommendations. We thank Carl E. McIntosh for his bioinformatics training. We thank the Laboratory of Molecular Evolution at the Hebrew University of Jerusalem for their continuous support throughout this project. This study was approved by the Hebrew University Agricultural Faculty's Animal Care and Use Committee under permit number AG-23476-07.

Supplemental Material

References

1. Aleman M, Gray LC, Williams DC, Holliday TA, Madigan JE, et al. (2006) Juvenile idiopathic epilepsy in Egyptian Arabian foals: 22 cases (1985-2005). J Vet Intern Med 20: 1443-9.

2. Lacombe VA, Mayes M, Mosseri S, Reed SM, Fenner WR, et al. (2012) Epilepsy in horses: Aetiological classification and predictive factors. Equine Vet J 44: 646-51.

3. Lichter-Peled A, Polani S, Stanyon R, Rocchi M, Kahila Bar-Gal G (2013) Role of KCNQ2 and KCNQ3 genes in juvenile idiopathic epilepsy in Arabian foals. Veterinary Journal 196: 57-63.

4. Mittel L (1987) Seizures in the horse. Vet Clin North Am Equine Pract 3: 323-32.

5. Fanelli HH (2005) Coat colour dilution lethal ("lavender foal syndrome"): a tetany syndrome of Arabian foals. Equine Vet Educ 17: 260-3.

6. Brooks SA, Gabreski N, Miller D, Brisbin A, Brown HE, et al. (2010) Whole-genome SNP association in the horse: Identification of a deletion in myosin Va responsible for Lavender Foal Syndrome. PLoS Genetics 6: 1000909.

7. Aleman M, Finno CJ, Weich K, Penedo MCT (2017) Investigation of Known Genetic Mutations of Arabian Horses in Egyptian Arabian Foals with Juvenile Idiopathic Epilepsy. J Vet Intern Med 32: 465-8.

8. Cook D, Gallagher PC, Bailey E (2010) Genetics of swayback in American Saddlebred horses. Anim Genet. 41: 64-71.

9. Fox-Clipsham LY, Carter SD, Goodhead I, Hall N, Knottenbelt DC, et al. (2011) Identification of a mutation associated with fatal foal immunodeficiency syndrome in the fell and dales pony. PLoS Genetics 7: doi: 10.1371.

10. Metzger J, Ohnesorge B, Distl O (2012) Genome-wide linkage and association analysis identifies major gene loci for guttural pouch tympany in Arabian and German warmblood horses. PLoS ONE 7: doi:10.1371.

11. Polani S (2017) The Genetics of Juvenile Idiopathic Epilepsy in Arabian Foals, The Hebrew University of Jerusalem, Israel.

12. Reich DE, Lander ES (2001) On the allelic spectrum of human disease. Trends Genet 17: 502-10.

13. Cirulli ET, Goldstein DB (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11: 415-25.

14. Kearney JA (2012) Advances in epilepsy genetics and genomics. Epilepsy Curr 12: 143-6.

15. Qin H, Zhu X (2012) Allowing for population stratification in association analysis. Methods Mol Biol 850: 399-409.

16. Bras J, Guerreiro R, Hardy J (2012) Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. Nat Rev Neurosci 13: 453-64.

17. Veeramah KR, O'Brien JE, Meisler MH, Cheng X, Dib-Hajj SD, et al. (2012) De novo pathogenic SCN8A mutation identified by whole-genome sequencing of a family quartet affected by infantile epileptic encephalopathy and SUDEP. Am J Hum Genet 90: 502-10.

18. McMullin D (2012) Epi4K: Gene discovery in 4,000 genomes. Epilepsia 53: 1457-67.

19. Do KT, Kong HS, Lee JH, Lee HK, Cho BW, et al. (2014) Genomic characterization of the Przewalski's horse inhabiting Mongolian steppe by whole genome re-sequencing. Livestock Science 167: 86-91.

20. Doan R, Cohen ND, Sawyer J, Ghaffari N, Johnson CD, et al. (2012) Whole-Genome sequencing and genetic variant analysis of a quarter Horse mare. BMC Genomics 13: 78.

21. Jun J, Cho YS, Hu H, Kim H-M, Jho S, et al. (2014) Whole genome sequence and analysis of the Marwari horse breed and its genetic origin. BMC Genomics 15: s4.

22. Lee JH, Song KD, Kim JM, Leem HK, Park KD (2016) Identification of genes with nonsynonymous SNP in Jeju horse by whole-genome resequencing reveals a functional role for immune response. J Anim Sci 94: 895-901.

23. Metzger J, Tonda R, Beltran S, Agueda L, Gut M, et al. (2014) Next generation sequencing gives an insight into the characteristics of highly selected breeds versus non-breed horses in the course of domestication. BMC Genomics 15: 562.

24. Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, et al. (2013) Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature 499: 74-8.

25. Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, et al. (2014) Prehistoric genomes reveal the genetic foundation and cost of horse domestication. Proc Natl Acad Sci U S A 111: e5661-9.

26. Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16: 1215.

27. Andrews S (2010) FastQC: A quality control tool for high throughput sequence data.

28. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, et al. (2009) Genome Sequence, Comparative Analysis and Population Genetics of the Domestic Horse. Science 326: 865-7.

29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-79.

30. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-60.

31. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43: 491-8.

32. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. (2010) The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20: 1297-1303.

33. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, et al. (2012) Qualimap: Evaluating next-generation sequencing alignment data. Bioinformatics 28: 2678-9.

34. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, et al. (2012) Control-FREEC: A tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 28: 423-5.

35. Derrien T, Estellé J, Sola SM, Knowles DG, Raineri E, et al. (2012) Fast computation and applications of genome mappability. PLoS ONE 7: doi:10.1371.

36. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w 1118; iso-2; iso-3. Fly 6: 80-92.

37. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, et al. (2011) Ensembl BioMarts: A hub for data retrieval across taxonomic space. Database 2011: doi:10.1093/database/bar030.

38. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-2.

39. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, et al. (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35: doi:10.1093/nar/gkm306.

40. Polani S, Roca AL, Rosensteel BB, Kolokotronis SO, Bar-Gal GK (2010) Evolutionary dynamics of endogenous feline leukemia virus proliferation among species of the domestic cat lineage. Virology 405: 397-407.

41. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, et al. (2011) The variant call format and VCFtools. Bioinformatics 27: 2156-8.

42. Herrero J, Muffato M, Beal K, Fitzgerald S, Gordon L, et al. (2016) Ensembl comparative genomics resources. Database (Oxford): doi:10.1093/database/bav096.

43. Choi Y, Chan AP (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31: 2745-7.

44. Mansour TA, Scott EY, Finno CJ, Bellone RR, Mienaltowski MJ, et al. (2017) Tissue resolved, gene structure refined equine transcriptome. BMC Genomics 18: doi:10.1186/s12864-016-3451-2.

45. Hestand MS, Kalbfleisch TS, Coleman SJ, Zeng Z, Liu J, et al. (2015) Annotation of the Protein Coding Regions of the Equine Genome. PLoS One 10: 1-13.

46. Pacholewska A, Drögemüller M, Klukowska-Rötzler J, Lanz S, Hamza E, et al. (2015) The Transcriptome of Equine Peripheral Blood Mononuclear Cells. PLoS One 10: 1-16.

47. Capomaccio S, Vitulo N, Verini-Supplizi A, Barcaccia G, Albiero A, et al. (2014) RNA Sequencing of the Exercise Transcriptome in Equine Athletes. PLoS One 8: 1-14.

48. Ropka-Molik K, Stefaniuk-Szmukier M, Żukowski K, Piórkowska K, Gurgul A, et al. (2017) Transcriptome profiling of Arabian horse blood during training regimens. BMC Genet 18: doi:10.1186/s12863-017-0499-1.

49. Doan R, Cohen N, Harrington J, Veazy K, Juras R, et al. (2012) Identification of copy number variants in horses. Genome Res 22: 899-907.

50. Ghosh S, Qu Z, Das PJ, Fang E, Juras R, et al. (2014) Copy Number Variation in the Horse Genome. PLoS Genet 10: doi:10.1371/journal.pgen.1004712.

51. Finno CJ, Bannasch DL (2014) Applied equine genetics. Equine Vet J 46: 538-44.

52. Zhang L, Mei Y, Fu N, Guan L, Xie W, et al. (2012) TRIM39 regulates cell cycle progression and DNA damage responses via stabilizing p21. Proc Natl Acad Sci U S A 109: 20937-42.

53. Hatakeyama S, Nakayama KI (2003) U-box proteins as a new family of ubiquitin ligases. Biochem Biophys Res Commun 302: 635-45.

54. Manchenkov T, Pasillas MP, Haddad GG, Imam FB (2015) Novel Genes Critical for Hypoxic Preconditioning in Zebrafish Are Regulators of Insulin and Glucose Metabolism. G3 (Bethesda, Md.) 5: 1107-16.

55. Meroni G, Diez-Roux G (2005) TRIM/RBCC, a novel class of "single protein RING finger" E3 ubiquitin ligases. Bioessays 27: 1147-57.

56. Ozato K, Shin D-M, Chang T-H, Morse HC (2008) TRIM family proteins and their emerging roles in innate immunity. Nat Rev Immunol 8: 849-60.

57. Kimura T, Mandell M, Deretic V (2016) Precision autophagy directed by receptor regulators - emerging examples within the TRIM family. J Cell Sci 129: 881-91.

58. Cheung CC, Yang C, Berger T, Zaugg K, Reilly P, et al. (2010) Identification of BERP (brain-expressed RING finger protein) as a p53 target gene that modulates seizure susceptibility through interacting with GABA(A) receptors. Proc Natl Acad Sci U S A 107: 11883-8.

59. Schreiber J, Végh MJ, Dawitz J, Kroon T, Loos M, et al. (2015) Ubiquitin ligase TRIM3 controls hippocampal plasticity and learning by regulating synaptic γ-actin levels. J Cell Biol 211: 569-86.

60. Sakai Y, Fukai R, Matsushita Y, Miyake N, Saitsu H, et al. (2016) De Novo Truncating Mutation of TRIM8 Causes Early-Onset Epileptic Encephalopathy. Ann Hum Genet 80: 235-40.

61. Wang Y, Li Z, Hu G, Lin L, Zeng L, et al. (2016) Molecular Characterization, Tissue Distribution and Expression, and Potential Antiviral Effects of TRIM32 in the Common Carp (Cyprinus carpio). Int J Mol Sci 17: 1693.

62. Romá-Mateo C, Moreno D, Vernia S, Rubio T, Bridges TM, et al. (2011) Lafora disease E3-ubiquitin ligase malin is related to TRIM32 at both the phylogenetic and functional level. BMC Evo Biol 11: 225.

63. Lionel AC, Crosbie J, Barbosa N, Goodale T, Thiruvahindrapuram B, et al. (2011) Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. Sci Transl Med 3: 95ra75.

64. Lionel AC, Tammimies K, Vaags AK, Rosenfeld JA, Ahn JW, et al. (2014) Disruption of the ASTN2/TRIM32 locus at 9q33.1 is a risk factor in males for autism spectrum disorders, ADHD and other neurodevelopmental phenotypes. Hum Mol Genet 23: 2752-68.

65. Balastik M, Ferraguti F, Pires-da Silva A, Lee TH, Alvarez-Bolado G, et al. (2008) Deficiency in ubiquitin ligase TRIM2 causes accumulation of neurofilament light chain and neurodegeneration. Proc Natl Acad Sci U S A 105: 12016-21.

66. Ylikallio E, Pöyhönen R, Zimon M, De Vriendt E, Hilander T, et al. (2013) Deficiency of the E3 ubiquitin ligase TRIM2 in early-onset axonal neuropathy. Hum Mol Genet 22: 2975-83.

67. Tanji K, Kamitani T, Mori F, Kakita A, Takahashi H, et al. (2010) TRIM9, a novel brain-specific E3 ubiquitin ligase, is repressed in the brain of Parkinson's disease and dementia with Lewy bodies. Neurobiol Dis 38: 210-18.

68. Shi M, Cho H, Inn K-S, Yang A, Zhao Z, et al. (2014) Negative regulation of NF-κB activity by brain-specific TRIpartite Motif protein 9. Nat Commun 5: 4820.

69. Niikura T, Hashimoto Y, Tajima H, Ishizaka M, Yamagishi Y, et al. (2003) A tripartite motif protein TRIM11 binds and destabilizes Humanin, a neuroprotective peptide against Alzheimer's disease-relevant insults. Eur J Neurosci 17: 1150-8.

70. Freudenberger P, Schmidt R, Schmidt H (2012) Genetics of age-related white matter lesions from linkage to genome wide association studies. J Neurol Sci 322: 82-6.

71. Kurata R, Nakaoka H, Tajima A, Hosomichi K, Shiina T, et al. (2010) TRIM39 and RNF39 are associated with Beh??et's disease independently of HLA-B*51 and -A*26. Biochem Biophys Res Commun 401: 533-7.

72. Kunz M, König IR, Schillert A, Kruppa J, Ziegler A, et al. (2015) Genome-wide association study identifies new susceptibility loci for cutaneous lupus erythematosus. Exp Dermatol 24: 510-5.

73. McDermott E, Ryan EJ, Tosetto M, Gibson D, Burrage J, et al. (2015) DNA methylation profiling in inflammatory bowel disease provides new insights into disease pathogenesis. Journal of Crohn's & Colitis 10: 77-86.

74. Suzuki M, Watanabe M, Nakamaru Y, Takagi D, Takahashi H, et al. (2016) TRIM39 negatively regulates the NF??B-mediated signaling pathway through stabilization of Cactin. Cell Mol Life Sci 73: 1085-101.

75. Tornatore L, Thotakura AK, Bennett J, Moretti M, Franzoso G (2012) The nuclear factor kappa B signaling pathway: Integrating metabolism with inflammation. Trends Cell Biol 22: 557-66.

76. O'Neill La, Kaltschmidt C (1997) NF-kappa B: a crucial transcription factor for glial and neuronal cell function. Trends Neurosci 20: 252-258.

77. Zhang Y, Hu W (2012) NF-kB signaling regulates embryonic and adult neurogenesis. Front Biol 7: 277-91.

78. Lubin FD, Ren Y, Xu X, Anderson AE (2007) Nuclear factor-κB regulates seizure threshold and gene transcription following convulsant stimulation. J Neurochem 103: 1381-95.

79. Di Q, Yu N, Liu H, Hu Y, Jiang Y, et al. (2011) Nuclear factor-kappa B activity regulates brain expression of P-glycoprotein in the kainic acid-induced seizure rats. Med Inflamm 2011: 670613.

80. Schinkel AH (1999) P-Glycoprotein, a gatekeeper in the blood-brain barrier. Adv Drug Deliv Rev 36: 179-94.

81. Blazyczek I, Hamann H, Ohnesorge B, Deegen E, Distl O (2004) Inheritance of guttural pouch tympany in the Arabian horse. J Hered 95: 195-9.

82. Brault LS, Famula TR, Penedo MCT (2011) Inheritance of cerebellar abiotrophy in Arabians. Am J Vet Res 72: 940-4.

83. Makvandi-Nejad S, Hoffman GE, Allen JJ, Chu E, Gu E, et al. (2012) Four loci explain 83% of size variation in the horse. PLoS ONE 7: e39929.

84. Biervert C, Schroeder BC, Kubisch C, Berkovic SF, Propping P, et al. (1998) A potassium channel mutation in neonatal human epilepsy. Science 279: 403-6.

85. Goldberg-Stern H, Kaufmann R, Kivity S, Afawi Z, Heron SE (2009) Novel Mutation in KCNQ2 Causing Benign Familial Neonatal Seizures. Pediatr Neurol 41: 367-70.

86. Leppert M, Anderson VE, Quattlebaum T, Stauffer D, O'Connell P, et al. (1989) Benign familial neonatal convulsions linked to genetic markers on chromosome 20. Nature 337: 647-8.

87. Singh NA, Westenskow P, Charlier C, Pappas C, Leslie J, et al. (2003) KCNQ2 and KCNQ3 potassium channel genes in benign familial neonatal convulsions: Expansion of the functional and mutation spectrum. Brain 126: 2726-37.

88. Charlier C, Singh NA, Ryan SG, Lewis TB, Reus BE, et al. (1998) A pore mutation in a novel KQT-like potassium channel gene in an idiopathic epilepsy family. Nat Genet 18: 53-5.

89. Concolino D, Iembo MA, Rossi E, Giglio S, Coppola G, et al. (2002) Familial pericentric inversion of chromosome 5 in a family with benign neonatal convulsions. J Med Genet 39: 214-6.

90. Davenport H (1913) The Arabian horse. J Hered 4: 76-8.

91. Edwards GB (1973) The Arabian: War Horse to Show Horse, Revised Co, Rich Publishing, Covina, California.

92. Lau AN, Peng L, Goto H, Chemnick L, Ryder OA, et al. (2009) Horse domestication and conservation genetics of Przewalski's horse inferred from sex chromosomal and autosomal sequences. Mol Biol Evol 26: 199-208.

93. Watson AG, Mayhew IG (1986) Familial congenital occipitoatlantoaxial malformation (OAAM) in the Arabian horse. Spine 11: 334-9.

94. Wiler R, Leber R, Moore BBB, VanDyk LFF, Perryman LEE, et al. (1995) Equine severe combined immunodeficiency: a defect in V(D)J recombination and DNA-dependent protein kinase activity. Proc Natl Acad Sci U S A 92: 11485-9.