Mammalian Genome · Springer

Canine genome assembly correction facilitates identification of a MAP9 deletion as a potential age of onset modifier for RPGRIP1 associated canine retinal degeneration

SCHOLARONE™
Manuscripts

Canine genome assembly correction facilitates identification of a *MAP9* deletion as a potential age of onset modifier for *RPGRIP1* associated canine retinal degeneration

Oliver P. Forman[1], Rebekkah J. Hitti[1*], Mike Boursnell[1], Keiko Miyadera[3], David Sargan[2], Cathryn Mellersh[1].

1. Kennel Club Genetics Centre, Animal Health Trust, Newmarket Suffolk, CB87UU. UK
2. Comparative Genetics Group, Department of Clinical Veterinary Medicine, University of Cambridge, Madingley Rd., Cambridge, CB3 0ES, UK
3. School of Veterinary Medicine, University of Pennsylvania, 3900 Delancey St, Philadelphia, PA 19104. USA

* corresponding author

Email: rebekkah.hitti@aht.org.uk
Telephone: +44 1638 750 659
Facsimile: +44 1638 555 666

Short Title: MAP9 associated with retinal degeneration onset age

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

Retinal degeneration (RD) in the Miniature Long Haired Dachshund (MLHD) is a cone-rod dystrophy resulting in eventual blindness in affected individuals. In a previous study a 44-nucleotide insertion (ins44) in exon 2 of *RPGRIP1* was associated with RD. However, results on an extended population of MLHD revealed a variable RD onset age for ins44 homozygous dogs. Further investigations using a genome-wide association study comparing early onset and late onset RD cases, identified an age of onset modifying locus for RD, approximately 30Mb upstream of *RPGRIP1* on chr15. In this investigation target enriched sequencing identified a *MAP9* deletion spanning approximately 22 kb associated with early RD onset. Identification of the deletion required correction to the CanFam3.1 genome build as canine *MAP9* is part of a historic tandem duplication, resulting in incomplete assembly of this genome region. The deletion breakpoints were identified in *MAP9* intron 10 and in a downstream partial *MAP9* pseudogene. The fusion of these two genes, which we have called *MAP9$_{EORD}$* (Microtubule associated protein, early onset retinal degeneration), is in frame and is expressed at the RNA level, with the 3' region containing several predicted deleterious variants. We speculate that *MAP9* associates with α-tubulin in the basal body of the cilium. RPGRIP1 is also known to locate to the cilium, where it is closely associated with RPGR. *RPGRIP1* mutations also cause redistribution of α-tubulin away from the ciliary region in photoreceptors. Hence a *MAP9* partial deficit is a particularly attractive candidate to synergise with a partial *RPGRIP1* deficit to cause a more serious disease.

Introduction

Retinal degeneration (RD) in the Miniature Long Haired Dachshund (MLHD), first described by Barnett et al (Curtis and Barnett 1993) and later further clinically characterised by Turney et al (Turney et al. 2007), is a cone-rod dystrophy where first the cone and then the rod photoreceptors of the retina become depleted resulting in eventual blindness. The disease was first characterised in a colony of MLHD segregating RD in a manner consistent with an autosomal recessive inheritance. Clinical signs included tapetal hyperreflectivity, early cone-specific ERG amplitude reduction, and, to a lesser extent, a reduction in rod ERG amplitude. Degeneration progressed fully with no detectable ERG responses after 40 weeks of age in RD affected colony dogs (Turney et al. 2007).

Genetic analysis of the RD which segregated within a closed breeding colony of MLHD identified a 44-nucleotide insertion (ins44) in exon 2 of the retinitis pigmentosa GTPase regulator-interacting protein 1 (*RPGRIP1*) gene (Mellersh et al. 2006). This variant was predicted to alter the reading frame, causing a frameshift and premature termination. The variant segregated perfectly within the breeding colony and given the strong association of *RPGRIP1* with Leber congenital amaurosis in human populations (Dryja et al. 2001; Gerber et al. 2001) and predictions for a truncated protein, the insertion was deemed likely to be the cause of RD for this particular dog breed.

Despite perfect concordance of ins44 with RD in the closed breeding colony, evidence from DNA testing laboratories, breeders and veterinarians indicated that there was considerable phenotype-genotype discordance within MLHDs that were not directly related to the experimental colony as not all dogs that were homozygous for ins44 were developing early onset RD (EORD), but in fact were developing RD at a much later stage or not at all within their lifespan. A study of 59 sporadic cases and 200 apparently normal controls from the MLHD pet dog population revealed 16 % of controls to be homozygous for ins44 and 20% of cases as non-homozygous for ins44 (Miyadera et al. 2009). In an independent study of 23 control MLHD and 34 ins44 homozygotes, it was demonstrated that the ins44 homozygous genotype does not invariably lead to the early cone-rod retinal dystrophy (cord) phenotype (Busse et al. 2011). Age of RD onset ranges of from 0.3 to 15 years among ins44 homozygous dogs in the wider population and subsequently an age of onset modifying locus was identified for RD in the MLHD using a genome-wide association study approach comparing EORD with late onset RD (LORD) cases, although the modifying variant was not identified (Miyadera et al. 2012b).

The ins44 mutation has also been identified in the English Springer Spaniel (ESS) and heterozygously in the Labrador Retriever and French bulldog dog breeds. In addition a slightly longer insert (ins69) was detected at the same position in Beagles (BE), including homozygous individuals (Miyadera et al. 2009). A small study investigating RD in ESS from the US showed that out of six ins44 homozygous ESS dogs, four showed clinical signs, suggesting that an age of onset modifying locus may also play a role in development of *RPGRIP1* associated disease in the ESS (Narfstrom et al. 2012). In the same study investigation of Swedish ESS showed eight out of ten cases of RD in the ESS were not homozygous for the ins44 variant suggesting potential alternative clinical or genetic causes of RD in this breed.

Since the identification of the *RPGRIP1* ins44 mutation, experimental data from retinal RNA studies of ins44 dogs has shown that rather than causing an insertion of an aberrant string of codons and a frameshift, the ins44 mutation actually results in an alternative splicing pattern and skipping of exon 2 of *RPGRIP1*. But despite exon 2 skipping in *cRPGRIP1Ins44/Ins44* dogs, the protein pattern detected in western blots with an antiRPGRIP1 antibody is the same as in *cRPGRIP1+/+* and

cRPGRIP1+/Ins44 retinal samples. (Kuznetsova et al. 2012). This study proposed exclusion of ins44 as the causal variant for RD and suggested an alternative (as yet unidentified) variant either within *RPGRIP1* or within the mapped region as the true cause of RD.

As *RPGRIP1* is associated with human Leber congenital amaurosis, *RPGRIP1* associated RD in the dog has been considered as a potential model for gene therapy studies since the identification of the ins44 variant (Lheriteau et al. 2009). In a recent study using ins44 homozygous dogs derived from the original colony presenting early onset RD cases, gene therapy was reported to have been successfully used to significantly and stably rescue cone function, by subretinal injection of adeno-associated virus (AAV) encoding canine *RPGRIP1* (Lheriteau et al. 2014). These experiments either provide evidence that a mutation in *RPGRIP1* or a mutation affecting its expression is causal of RD in the dog, or that overexpression of *RPGRIP1* is capable of re-establishing a normal phenotype.

Since its identification, the ins44 insertion has been subject to scientific investigation, discussion and debate. However, regardless of whether the ins44 mutation is the causal mutation for RD in the MLHD, or whether the causal mutation is an alternative variant in close linkage disequilibrium with the ins44 variant, there is clear evidence that the age of onset of RPGRIP1-associated RD in this breed is influenced by a modifying locus, approximately 30Mb away on chr15 (Miyadera et al. 2012b) The purpose of this study was to identify the age modifying variant, with the aim of improving the understanding of *RPGRIP1* associated RD.

Materials and methods

DNA was extracted from whole blood using the Nucleon BACC2 DNA extraction kit (GE Healthcare). DNA was extracted from buccal swabs using the QIAamp Midi kit (Qiagen).

Probes for target enrichment were designed using the design tool eArray with libraries prepared and sequenced at The Genome Analysis Centre, Norwich. The library for genome sequencing was prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina, and sequencing performed on the Illumina MiSeq platform, generating an 82.5Gb dataset of 300 bp paired-end reads. Alignment of reads to the reference genome CanFam3.1 was performed using BWA (Li and Durbin 2009) and variants were called using GATK (McKenna et al. 2010).

Screening of 23,040 clones of a canine BAC library (RPCI-81: Male Doberman Pinscher) was performed by qPCR of diluted cell lysates of glycerol stocks. Quantitative PCRs were carried out using KAPA Probe Fast qPCR ready mix on an ABI stepone plus real-time PCR machine. Positive BAC clones were streaked from glycerol stocks onto LB agar containing 12.5 mg/ml chloramphenicol, and picked colonies grown in liquid LB containing 12.5 mg/ml chloramphenicol. BAC DNA was isolated using the ZR BAC DNA Miniprep Kit. Long PCRs were performed using KAPA HIFI Hot Start Ready Mix. Libraries for sequencing of BAC DNA and long PCR products on the Illumina MiSeq platform were prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina. *De novo* assembly was performed using SOAPdenovo (Xie et al. 2014).

RNA was extracted from spleen using the Qiagen RNeasy Midi Kit. Reverse transcription was performed using the Qiagen Quantitect reverse transcription kit. Expression analysis of *MAP9* transcripts was performed using KAPA Probe Fast qPCR ready mix on an Illumina Eco qPCR machine. All primers are listed in Supplementary File 1.

Results

Target enriched massively-parallel sequencing was used to generate sequence data for two EORD cases and eight controls (LORD cases), all of which were homozygous for the *RPGRIP1* ins44 variant. The genomic target was the mapped interval for the age of onset modifying locus (chr15:51,980,056-53,979,789 CanFam3.1), with 82 % of the target region covered by at least two reads and 73 % of the target region achieved at least 20x coverage.

Sequence read alignments to the canine genome build CanFam3.1 identified 11,494 SNP and indel variants. Simple variants were filtered by how closely they segregated with disease status and consequence predictions. After filtering only one plausible variant remained – a SNP causing an amino acid substitution (S107Y) in the gene encoding cathepsin O (*CTSO*), which was predicted to be damaging through computational analysis using SIFT (Ng and Henikoff 2003). By coincidence, this variant is included on the Illumina CanineSNP20 bead chip and had thus been genotyped in the cohort of MLHDs investigated by the genome-wide association study to map the modifying locus (ID: BICF2S2353161). Despite being the single SNP with the lowest associated p-value (top SNP), the presence of several late onset individuals that were homozygous for the disease-associated *CTSO* variant meant that this variant could be excluded.

Further investigation of the modifying locus was carried out by visualising read alignments exon by exon and considering differences between EORD cases and controls (LORD cases). A region of zero read depth in EORD cases, but substantial read depth in controls was identified across exon 9 of the gene encoding microtubule associated protein 9 (*MAP9*) (Figure 1). The presence of a large number of reads aligning with a mapping quality of zero, which indicate more than one possible genomic coordinate for an alignment, suggested that the region was repetitive in nature. In addition no reads were identified that spanned the region of zero read depth for early onset cases, which would usually indicate a deletion. The reads with a mapping quality of zero also aligned perfectly to chrUn_AAEX03022082, a 21 kb scaffold with no designated chromosome. A gene with high sequence similarity to *MAP9* exon 8 through to exon 14, including introns, was present on chrUn_AAEX03022082. Short PCR across the region of zero read depth yielded products with no size difference between EORD and LORD cases due to the level of sequence similarity between the *MAP9* region on chr15 and the *MAP9-like* region on chrUn_AAEX03022082. Although no capture probes were specifically designed to capture chrUn_AAEX03022082, the region was inadvertently captured due to the high levels of sequence similarity. In addition to the zero read depth shown for exon 9 of *MAP9* on chr15 for EORD cases, no read alignments were seen for exons 11 through to 14 of the *MAP9*-like gene on chrUn_AAEX03022082. We sought to assemble chrUn_AAEX03022082 into the genome sequence to help determine the exact nature of the sequence variant identified.

In an attempt to determine the exact genomic assembly point for chrUn_AAEX03022082, screening of a canine BAC library (RPCI-81: Male Doberman Pinscher) was carried out to isolate clones containing chrUn_AAEX03022082. Two clones were identified, 36I10 and 40O5, for which BAC DNA was sequenced using a massively parallel approach. BAC ends were identified at coordinates chr15:52,766,756 and chr15:52,920,909 for isolate 40O5 indicating an assembly point within the chr15 *MAP9* region. Sequencing data alignments for BAC isolate 36I10 identified BAC ends at chr15:52,778,639 and chrUn_AAEX03022082:3,729, identifying the chromosome 15 assembly point. *De novo* assembly of the BAC sequencing data was used to bridge the 5' assembly point between chr15 and ChrUn_AAEX03022082.

Although the 5' assembly position had been determined through BAC sequencing further PCR based investigation was required to determine the 3' assembly point. Alignment of chrUn_AAEX03022082 to chr15 using BLAST identified two regions only present on chromosome 15 that had no similarity

to chrUn_AAEX03022082, allowing specific long PCR spanning from chr15 into chrUn_AAEX03022082 across the 3' assembly point (Figure 2). Sequencing and *de novo* assembly of the long PCR product enabled the new assembly of chromosome 15 containing chrUn_AAEX03022082 to be completed. The assembly point for chrUn_AAEX03022082 was chr15:52,909,237 with a new end point of 52,930,899. Bases 1-74 of chrUn_AAEX03022082 (part of *MAP9* exon 8) were corrected as part of the assembly process. A novel 280 bp sequence was added after chrUn_AAEX03022082 (chr15:52,930,899-52,931,179). We have given the new genome assembly the nominal title of CanFam3.1$_{MAP9\_corrected}$.

The new assembly shows that the *MAP9-like* exons, which were previously part of chrUn_AAEX03022082, actually form the 3' region of the *MAP9* gene. The 3' region of *MAP9* on the previous assembly (exons 8 to 14) is the result of a partial duplication of the *MAP9* gene and is a *MAP9* pseudogene. Review of the BLAST results suggest exon 6 and 7 were part of the duplicated region, but have subsequently been deleted.

Alignment of genome data generated for an EORD case and target enrichment data from EORD cases and controls to the newly assembled CanFam3.1$_{MAP9\_corrected}$, were suggestive of a deletion in EORD cases potentially through homologous recombination. The deletion was confirmed in EORD cases and not LORD cases by long PCR to produce a product spanning the deletion joins. A PCR specific to the deleted region yielded products only for LORD cases (Figure 3). Definition of the deletion breakpoints could only be defined to be within a 1666 bp region through visualising alignments of targeted enrichment and genome data due to low read mapping quality and spurious alignments (Supplementary File 2). Although precise and certain definition of the deletion breakpoint was impossible due to the level of sequence similarity between the two genes, a minimal deletion join/breakpoint region of 231 bp was identified using multiple sequence alignment of *de novo* assembled reads generated from the PCR product spanning the deletion (Supplementary File 3). The 5' region of the join showed sequence similarity to the region upstream of the 5' deletion breakpoint and the 3' region of the join showed sequence similarity to region downstream of the 3' deletion breakpoint. The size of deletion based on new genome build CanFam3.1$_{MAP9\_corrected}$ is 21,961 bp, with deletion breakpoints in intron 10 of *MAP9* and *MAP9* pseudogene(chr15:52,905,335-52,905,565 and chr15:52,927,296-52,927,527 (chrUn_AAEX03022082:3,379-3,609)). The new genome arrangement with key coordinates are shown graphically in Figure 4 and a timeline of chromosome 15 rearrangement events summarised in Figure 5.

Despite the deletion, the joining of *MAP9* with the *MAP9like* pseudogene appears to produce a complete *MAP9* reading frame which we have termed *MAP9$_{EORD}$*. Sequencing of cDNA from spleen of an EORD case (used as retina tissue was unavailable) confirmed that a full length product and in frame transcript was produced from the *MAP9$_{EORD}$* gene. A number of residue changes were identified between the canine *MAP9$_{EORD}$* and canine *MAP9* transcripts. Multiple sequence alignment was performed alongside human *MAP9*, and vertebrate species conservation was considered for variant residues (Supplementary file 4). The PolyPhen tool was used to predict whether the variants could alter protein function. The variant N555K was private to EORD cases on consideration of vertebrate species alignments, and was predicted by PolyPhen to be probably deleterious. The variant N579S was also predicted by PolyPhen to be probably deleterious and is a highly conserved residue across species. The deleted residue E584 is also highly conserved amongst vertebrate species.

Association of *MAP9$_{EORD}$* with the EORD phenotype was investigated by genotyping a subset of the EORD and LORD cases used for mapping of the early onset locus. The arbitrary age boundary for EORD cases was set at 5 years of age, and LORD cases were defined as diagnosed with RD at >5

years of age. In total there were 32 EORD cases and 18 LORD cases. Of the 30 EORD cases that genotyped successfully 26 were homozygous for $MAP9_{EORD}$ and for the LORD cases, one was homozygous for $MAP9_{EORD}$ (Table 1). The LORD case, which was homozygous for $MAP9_{EORD}$ had an age of diagnosis of 5.2 years of age i.e. close to the arbitrary cutoff, and quite likely to have had an onset age below 5y. In addition the sample set was genotyped for the top SNP from the age of onset locus genome-wide association study. Allelic association analysis performed on the top SNP and the $MAP9_{EORD}$ genotypes gave p-values of 1.4×10^{-5} and 4.0×10^{-8} respectively. A full description of results is show in Supplementary File 5.

The gene duplication, which is present in the canine reference sequence (Boxer) and the MLHD, was investigated in 96 dogs of 32 other breeds , by specific qPCR across the duplication join (i.e. within the *MAP9* deletion region). Three dogs of each of the following breeds were investigated: American Cocker Spaniel, Australian Shepherd, Boxer, Border Collie, Brittany Spaniel, Chesapeake Bay retriever, Cocker Spaniel, Doberman, English Springer Spaniel, Flat Coated Retriever, Field Spaniel, Grand Basset Griffon Vendeen, Great Dane, Golden Retriever, Giant Schnauzer, Hovawarts, Irish setter, Italian Spinone, Lancashire Heeler, Labrador Retriever, Miniature Bull Terrier, Miniature Poodle, Otterhound, Staffordshire Bull Terrier, Scottish Deerhounds, Soft-coated Wheaten Terrier, Standard Poodle, Shetland Sheepdogs, Tibetan Spaniel, Tibetan Terrier, West Highland White Terrier, and Welsh Springer Spaniel. Deletion alleles were identified in two individuals of two breeds, the Tibetan Terrier and Brittany. These alleles were further investigated using long PCR followed by massively parallel sequencing. Read alignment patterns and SNP analysis suggested a novel deletion event had occurred in the Tibetan Spaniel. The Brittany deletion showed high similarity to the deletion identified for the MLHD (Supplementary File 6). Genome comparison by BLAST suggests that the duplication is not present in the cat, the phylogenetically most similar genome sequence to the dog available.

To determine the frequency of the $MAP9_{EORD}$ allele within the Miniature Dachshund (MD) and ESS populations, 640 MDs (202 long haired, 348 smooth haired and 90 wire haired) and 206 ESSs were genotyped. Two MLHDs were homozygous for $MAP9_{EORD}$ (one ins44 homozygote and one ins44 heterozygote). A carrier frequency of 17.92% for the MLHD population was calculated assuming Hardy Weinberg equilibrium (q= 0.0995). No $MAP9_{EORD}$ homozygotes were identified in the ESSs sample set.

Discussion

Cone-rod RD in the MLHD is an example of a canine disease that segregates as a seemingly early onset recessive trait in an inbred population but manifests differently within a wider population albeit of the same breed. In this investigation we built on a previous age of onset locus mapping study, by using target enrichment data to identify a genomic deletion across *MAP9* as the potential age of onset modifying mutation.

The primary defect in the cone rod degeneration in a colony of MLHDs is known to be in *RPGRIP1*, caused by either or both of a polyA tract insertion in exon 2 of the gene (Mellersh et al. 2006) and a mutation elsewhere in the same confidence interval (Kuznetsova et al. 2012). The *RPGRIP1* mutation is leaky (Miyadera et al. 2012a; Miyadera et al. 2009) and gives rise to LORD or no clinically significant phenotype in the absence of a second homozygous mutation segregating in the interval containing *MAP9* (Miyadera et al. 2012b). Here we have shown that *RPGRIP1* ins44/ins44 dogs with EORD have several polymorphisms in *MAP9*, some of them potentially harmful, when compared with *MAP9* in LORD dogs. We hypothesise that this reflects a partial deficit in *MAP9* protein function in

the retina. This finding adds to a growing number of examples of modifiers affecting the RD phenotype mostly in mice models as well as in humans (Khanna et al. 2009; Maddox et al. 2012; Markand et al. 2015; Schon et al. 2016; Venturini et al. 2012; Vollrath et al. 2015).

MAP9, also known as ASAP (aster-associated protein), is associated with microtubules in interphase, with the mitotic spindle at mitosis, and with the centrosome during cytokinesis. It binds directly to pre-polymerised α-tubulin microtubules in vitro through its C-terminal domain, and co-sediments with cellular microtubules ex-vivo (Saffin et al. 2005). Phosphorylation of MAP9 at serine 625 (human, 626 in dog) by Aurora A kinase is essential for correct assembly onto centrosomes and for mitotic progression (Venoux et al. 2008). This residue is preserved in both the MAP9 proteins described here. MAP9 is probably expressed in all cell types, but shows high level expression in testes, brain, eye, spinal cord, fertilised ovum, and intermediate expression in kidney, heart, lung and a small number of other tissues. Within the eye, expression is known to be high in the retina, at least in the zebrafish (Saffin et al. 2005). Expression has not been mapped at this level in other species. In non-dividing cells such as photoreceptors the centrosome contributes a centriole to form the basal body of the primary cilium. Although not proven, it is likely that MAP9 remains associated with the α−tubulin in the basal body and MAP9 has been associated with the cytoskeletal component of the murine photoreceptor cilium (Liu et al. 2007), whilst in zebra fish, eye development is especially sensitive to morpholino driven reduction in MAP9 availability (Fontenille et al. 2014). RPGRIP1 is also known to locate to the ciliary axoneme, where it is closely associated with RPGR, and appears to act as a ciliary targeting protein for RPGR, NPHP4 and SDCCAG8 (Bolan and Wright, 2000; Hong et al. 2001; Patil et al. 2012; Roepman et al. 2005; Shu et al., 2005; Zhao et al. 2003) mutations in each of which cause a different retinal degeneration in dogs or other species. These degenerations have all been commonly described as ciliopathies. *RPGRIP1* mutations have profound effects on the stability and maintenance of the connecting cilium and its ability to support outer segment morphogenesis (e.g. Zhao et al., 2003). Changes in the cilium in RPGRIP1⁻ animals include reductions in acetylated α−tubulin in the axoneme in photoreceptors (Patil et al. 2012). Hence a MAP9 partial deficit is a particularly attractive candidate to synergise with a partial RPGRIP1 deficit to cause a more serious disease. It is notable that MAP9 is highly expressed in spermatogonia and in spermatocytes (Saffin et al. 2005). It is tempting to speculate that the original colony of cord1 (EORD) MLHD, in which low fertility was a persistent problem (Debra Flack, personal communication) and the EORD haplotype at *MAP9* was fixed, may have been showing a syndromic effect of the *MAP9* mutation described here. On the other hand up to one third of all retinopathies are non-syndromic ciliary defects even though many of the mutated genes are expressed widely outside the eye (Estrada-Cuzcano et al. 2012).

Identification of the *MAP9* deletion was facilitated by correction of the canine genome build CanFam3.1. In the new assembly, which we have given the nominal title CanFam3.1$_{MAP9\_corrected,}$ there is evidence of a historic partial duplication of *MAP9* through a hypothesised homologous recombination event. The tandem repeat structure explains the difficulty in assembling this region for the CanFam3.1 genome build. A subsequence deletion event, which again is likely to have occurred through homologous recombination, gives rise to a fusion of two halves of the partially duplicated *MAP9* gene to give a new gene, *MAP9$_{EORD}$*, which is strongly associated with the EORD phenotype. Investigation of other breeds for this variant identified individuals of two breeds, the Tibetan Terrier and Brittany Spaniel, with a similar gene arrangement to *MAP$_{EORD}$*. Haplotype analysis of these breeds showed that at least one of the deletion events appeared novel, further highlighting the region's vulnerability to rearrangement (Supplementary File 6). From our available clinical data there appears to be no phenotype associated solely with *MAP9$_{EORD}$* although formal investigation would be required to rule this possibility out.

Although the CanFam3.1 genome build is largely complete, there are still 15 unplaced scaffolds (unattached scaffolds, which are not assigned to a chromosomal location) and 223 unlocalised scaffolds, for which the chromosome is known, but the position and orientation is not. As many of these scaffolds are likely to contain very repetitive regions of the genome, it is likely that further improvement to the genome build will require the use of ultra-long read sequencing technology. Analysing short massively parallel sequence read alignments across repetitive regions is challenging. Although sequence reads may be accurate, a single variant in the reference sequence can result in reads aligning to an alternative region, leading to a gap in coverage and potentially falsely indicating a deletion. The opposite may also be also be true, where a deletion may be present but if reads map equally to both the deleted region and an alternative region, then the deleted region could be masked to some extent by reads of mapping quality zero. Both of these factors made detection of the deletion region and determination of the likely deletion breakpoints challenging.

Diagnostic DNA testing, which is commonly used in the dog, for $MAP9_{EORD}$ is challenging as a short PCR based assay is difficult to develop due to the unavailability of specific priming sites and long range PCR techniques may lack sensitivity, especially for DNA samples extracted from buccal swabs. Detection of presence or absence of MAP_{EORD} by qPCR can be used to specify early onset or late onset status for ins44 homozygotes however, and also enables MAP_{EORD} allele frequency estimates. Our data suggests $MAP9_{EORD}$ frequency is low in the MD population, implying that the majority of *RPGRIP1* ins44 dogs are not likely to have the EORD phenotype. No $MAP9_{EORD}$ alleles were detected in the ESS sample set. Although a larger sample set would be required to rule out the presence of $MAP9_{EORD}$, the result suggests that early onset retinal degeneration in the ESS is due to an alternative modifier or a separate genetic or clinical cause.

To summarise we have identified a deletion across the tandemly duplicated *MAP9* gene in the canine genome which segregates closely with EORD in the MLHD. Although a full *MAP9* open reading frame is maintained, the deletion results in the introduction of a number of variants which are predicted to be deleterious. This is a consequence of the 3' region being formed from a redundant *MAP9* partial pseudogene. We speculate that the interaction between *MAP9* and *RPGRIP1* are important in the function of the retina, and disruption of *MAP9* in the context of *RPGRIP1* ins44 mutation homozygosity result in the EORD phenotype.

Using different inbred populations of dog we have been able to tease apart the genetic basis of an inherited trait with a variable age of onset and demonstrate the interaction of two independent loci that contribute to the phenotype. The variation in age of onset between populations is explained by the fact that one of the populations was fixed for the modifier locus so the trait effectively segregated as a simple Mendelian trait in that population, an important consideration for the study of recessive conditions with small sample sets from closed or restricted populations

References

Busse C, Barnett KC, Mellersh CS, Adams VJ (2011) Ophthalmic and cone derived electrodiagnostic findings in outbred Miniature Long-haired Dachshunds homozygous for a RPGRIP1 mutation. Veterinary ophthalmology 14, 146-152

Boylan JP, Wright AF. Identification of a novel protein interacting with RPGR. Hum Mol Genet. 2000;**9**(14):2085-93. PubMed PMID: 10958647.

Curtis R, Barnett KC (1993) Progressive retinal atrophy in miniature longhaired dachshund dogs. The British veterinary journal 149, 71-85

Dryja TP, Adams SM, Grimsby JL, McGee TL, Hong DH, Li T, Andreasson S, Berson EL (2001) Null RPGRIP1 alleles in patients with Leber congenital amaurosis. Am J Hum Genet 68, 1295-1298

Estrada-Cuzcano A, Roepman R, Cremers FP, den Hollander AI, Mans DA (2012) Non-syndromic retinal ciliopathies: translating gene discovery into therapy. Hum Mol Genet 21, R111-124

Fontenille L, Rouquier S, Lutfalla G, Giorgi D (2014) Microtubule-associated protein 9 (Map9/Asap) is required for the early steps of zebrafish development. Cell Cycle 13, 1101-1114

Gerber S, Perrault I, Hanein S, Barbet F, Ducroq D, Ghazi I, Martin-Coignard D, Leowski C, Homfray T, Dufier JL, Munnich A, Kaplan J, Rozet JM (2001) Complete exon-intron structure of the RPGR-interacting protein (RPGRIP1) gene allows the identification of mutations underlying Leber congenital amaurosis. Eur J Hum Genet 9, 561-571

Hong D-H, Yue G, Adamian M and Li T. Retinitis Pigmentosa GTPase Regulator (RPGR)-interacting Protein Is Stably Associated with the Photoreceptor Ciliary Axoneme and Anchors RPGR to the Connecting Cilium. The Journal of Biological Chemistry 2001; **276**: 12091-12099.Khanna H, Davis EE, Murga-Zamalloa CA, Estrada-Cuzcano A, Lopez I, den Hollander AI, Zonneveld MN, Othman MI, Waseem N, Chakarova CF, Maubaret C, Diaz-Font A, MacDonald I, Muzny DM, Wheeler DA, Morgan M, Lewis LR, Logan CV, Tan PL, Beer MA, Inglehearn CF, Lewis RA, Jacobson SG, Bergmann C, Beales PL, Attie-Bitach T, Johnson CA, Otto EA, Bhattacharya SS, Hildebrandt F, Gibbs RA, Koenekoop RK, Swaroop A, Katsanis N (2009) A common allele in RPGRIP1L is a modifier of retinal degeneration in ciliopathies. Nat Genet 41, 739-745

Kuznetsova T, Iwabe S, Boesze-Battaglia K, Pearce-Kelling S, Chang-Min Y, McDaid K, Miyadera K, Komaromy A, Aguirre GD (2012) Exclusion of RPGRIP1 ins44 from primary causal association with early-onset cone-rod dystrophy in dogs. Invest Ophthalmol Vis Sci 53, 5486-5501

Lheriteau E, Libeau L, Stieger K, Deschamps JY, Mendes-Madeira A, Provost N, Lemoine F, Mellersh C, Ellinwood NM, Cherel Y, Moullier P, Rolling F (2009) The RPGRIP1-deficient dog, a promising canine model for gene therapy. Mol Vis 15, 349-361

Lheriteau E, Petit L, Weber M, Le Meur G, Deschamps JY, Libeau L, Mendes-Madeira A, Guihal C, Francois A, Guyon R, Provost N, Lemoine F, Papal S, El-Amraoui A, Colle MA, Moullier P, Rolling F (2014) Successful gene therapy in the RPGRIP1-deficient dog: a large model of cone-rod dystrophy. Molecular therapy : the journal of the American Society of Gene Therapy 22, 265-277

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760

Liu Q, Tan G, Levenkova N, Li T, Pugh EN, Jr., Rux JJ, Speicher DW, Pierce EA (2007) The proteome of the mouse photoreceptor sensory cilium complex. Molecular & cellular proteomics : MCP 6, 1299-1317

Maddox DM, Ikeda S, Ikeda A, Zhang W, Krebs MP, Nishina PM, Naggert JK (2012) An allele of microtubule-associated protein 1A (Mtap1a) reduces photoreceptor degeneration in Tulp1 and Tub Mutant Mice. Invest Ophthalmol Vis Sci 53, 1663-1669

Markand S, Saul A, Tawfik A, Cui X, Rozen R, Smith SB (2015) Mthfr as a modifier of the retinal phenotype of Crb1 mice. Experimental eye research 145, 164-172

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297-1303

Mellersh CS, Boursnell ME, Pettitt L, Ryder EJ, Holmes NG, Grafham D, Forman OP, Sampson J, Barnett KC, Blanton S, Binns MM, Vaudin M (2006) Canine RPGRIP1 mutation establishes cone-rod dystrophy in miniature longhaired dachshunds as a homologue of human Leber congenital amaurosis. Genomics 88, 293-301

Miyadera K, Brierley I, Aguirre-Hernandez J, Mellersh CS, Sargan DR (2012a) Multiple mechanisms contribute to leakiness of a frameshift mutation in canine cone-rod dystrophy. PLoS One 7, e51598

Miyadera K, Kato K, Aguirre-Hernandez J, Tokuriki T, Morimoto K, Busse C, Barnett K, Holmes N, Ogawa H, Sasaki N, Mellersh CS, Sargan DR (2009) Phenotypic variation and genotype-phenotype discordance in canine cone-rod dystrophy with an RPGRIP1 mutation. Mol Vis 15, 2287-2305

Miyadera K, Kato K, Boursnell M, Mellersh CS, Sargan DR (2012b) Genome-wide association study in RPGRIP1(-/-) dogs identifies a modifier locus that determines the onset of retinal degeneration. Mamm Genome 23, 212-223

Narfstrom K, Jeong M, Hyman J, Madsen RW, Bergstrom TF (2012) Assessment of hereditary retinal degeneration in the English springer spaniel dog and disease relationship to an RPGRIP1 mutation. Stem cells international 2012, 685901

Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31, 3812-3814

Patil H, Tserentsoodol N, Saha A, Hao Y, Webb M, Ferreira PA (2012) Selective loss of RPGRIP1-dependent ciliary targeting of NPHP4, RPGR and SDCCAG8 underlies the degeneration of photoreceptor neurons. Cell death & disease 3, e355

Roepman R, Letteboer SJ, Arts HH, van Beersum SE, Lu X, Krieger E, Ferreira PA, Cremers FP (2005) Interaction of nephrocystin-4 and RPGRIP1 is disrupted by nephronophthisis or Leber congenital amaurosis-associated mutations. Proc Natl Acad Sci U S A 102, 18520-18525

Saffin JM, Venoux M, Prigent C, Espeut J, Poulat F, Giorgi D, Abrieu A, Rouquier S (2005) ASAP, a human microtubule-associated protein required for bipolar spindle assembly and cytokinesis. Proc Natl Acad Sci U S A 102, 11302-11307

Schon C, Asteriti S, Koch S, Sothilingam V, Garcia Garrido M, Tanimoto N, Herms J, Seeliger MW, Cangiano L, Biel M, Michalakis S (2016) Loss of HCN1 enhances disease progression in mouse models of CNG channel-linked retinitis pigmentosa and achromatopsia. Hum Mol Genet

Shu X, Fry AM, Tulloch B, Manson FD, Crabb JW, Khanna H, Faragher AJ, Lennon A, He S, Trojan P, Giessl A, Wolfrum U, Vervoort R, Swaroop A, Wright AF. RPGR ORF15 isoform co-localizes with RPGRIP1 at centrioles and basal bodies and interacts with nucleophosmin. Hum Mol Genet. 2005 May 1;14(9):1183-97.

Turney C, Chong NH, Alexander RA, Hogg CR, Fleming L, Flack D, Barnett KC, Bird AC, Holder GE, Luthert PJ (2007) Pathological and electrophysiological features of a canine cone-rod dystrophy in the miniature longhaired dachshund. Investigative ophthalmology & visual science 48, 4240-4249

Venoux M, Basbous J, Berthenet C, Prigent C, Fernandez A, Lamb NJ, Rouquier S (2008) ASAP is a novel substrate of the oncogenic mitotic kinase Aurora-A: phosphorylation on Ser625 is essential to spindle formation and mitosis. Hum Mol Genet 17, 215-224

Venturini G, Rose AM, Shah AZ, Bhattacharya SS, Rivolta C (2012) CNOT3 is a modifier of PRPF31 mutations in retinitis pigmentosa with incomplete penetrance. PLoS Genet 8, e1003040

Vollrath D, Yasumura D, Benchorin G, Matthes MT, Feng W, Nguyen NM, Sedano CD, Calton MA, LaVail MM (2015) Tyro3 Modulates Mertk-Associated Retinal Degeneration. PLoS Genet 11, e1005723

Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, Zhou X, Lam TW, Li Y, Xu X, Wong GK, Wang J (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30, 1660-1666

Zhao Y, Hong DH, Pawlyk B, Yue G, Adamian M, Grynberg M, Godzik A, Li T (2003) The retinitis pigmentosa GTPase regulator (RPGR)- interacting protein: subserving RPGR function and participating in disk morphogenesis. Proc Natl Acad Sci U S A 100, 3965-3970

Table 1: Genotype phenotype correlation for *MAP9$_{EORD}$*

MAP9$_{EORD}$ genotype	wt/wt	wt/del	del/del
EORD cases	1	1	21
LORD cases	13	9	1

Figure 1: Area of zero read depth in target enrichment sequence data for EORD cases.

Alignment of target enrichment sequence read data against the CanFam3.1 reference sequence revealed a region of zero read depth for EORD cases, but not for controls (LORD cases). The presence of a large number of reads in the region with zero mapping quality, suggested a region(s) of very high sequence similarity existed elsewhere in the genome. No paired alignments with greater than expected insert size were present, usually indicative of a deletion.

Figure 2: BLAST results of chrUn_AAEX03022082 to chromosome 15.

Alignment of the chrUn_AAEX03022082 sequence (A) against chromosome 15 (B) using BLAST, revealed two regions which were specific to chromosome 15 (marked with asterisks). Corresponding alignment blocks are number from 1-12 and coloured from red through to blue respectively. An apparent sequence join was identified on chrUn_AAEX03022082 (marked with a green dashed line), suggestive of a structural rearrangement event.

Figure 3: *MAP9$_{EORD}$* genotyping PCR

Genotyping of *MAP9$_{EORD}$* was carried out by performing one long PCR across the deletion breakpoints and a specific PCR within the deleted region. PCR products were pooled before running on an agarose gel. Lanes 2-4 show a single band representing a PCR product crossing the deletion join (All EORD cases). Lanes 5-6 show a single band representing the PCR within the deleted region. Lane 7 is a heterozygous individual.

Figure 4: New assembly arrangement of chromosome 15.

The newly assembled chromosome 15 sequence contains the *MAP9* gene and a partial *MAP9* pseudogene which we have termed *MAP9-like*. The arrangement is likely to have occurred through partial tandem duplication of *MAP9* caused by homologous recombination.

Figure 5: Overview of events leading to _MAP9__{EORD}

(A) Homologous recombination event leading to creation of a partial MAP9like pseudogene. This pseudogene is present in all 32 dog breeds investigated for this study, but may be private to the dog, although other closely related canid species have not been investigated. (B) Accumulation of variants in the non-conserved _MAP9-like_ sequence. (C) Homologous recombination event leading to the creation of _MAP9__{EORD}. Although this gene is complete and expressed it possesses many variants compared to wildtype _MAP9,_ that change conserved residues of the _MAP9_ protein.

Supplementary File 1: Primer sequences

Supplementary File 2: Sequence read alignments across the deletion breakpoint regions.

Definition of the deletion breakpoints was not aided by alignment of targeted enrichment and genome data as poor mapping quality and spurious alignments meant the deletion breakpoint could only be defined to be within a 1666 bp region.

Supplementary File 3: Multiple sequence alignments

Multiple sequencing alignments of the de novo assembly of reads generated from the PCR product spanning the deletion to determine the most likely location of the deletion breakpoints.

Supplementary File 4: Vertebrate species conservation for _MAP9__{EORD} variant residues.

Supplementary File 5: Top SNP and _MAP9__{EORD} genotyping results.

Supplementary File 6: Alignment of massively parallel sequencing reads generated for a Tibetan Terrier and Brittany Spaniel with a deletion across _MAP9_
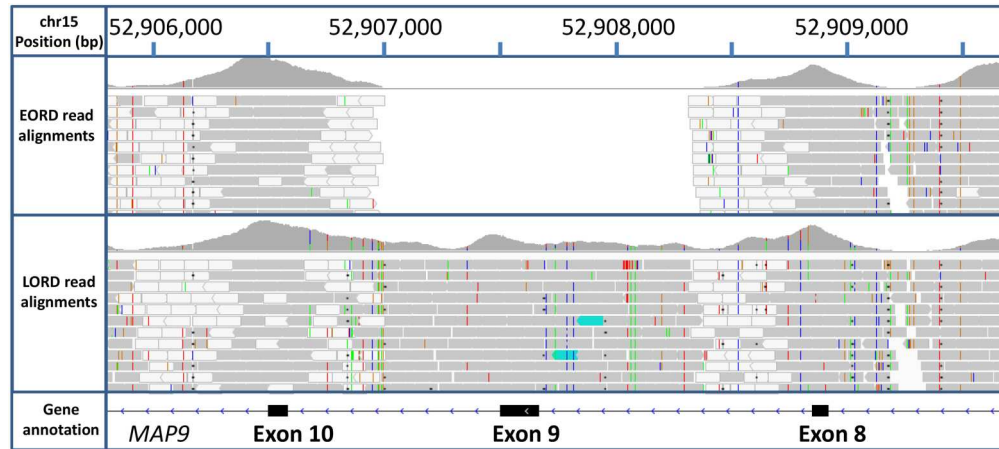
Figure 1: Area of zero read depth in target enrichment sequence data for EORD cases.
Alignment of target enrichment sequence read data against the CanFam3.1 reference sequence revealed a region of zero read depth for EORD cases, but not for controls (LORD cases). The presence of a large number of reads in the region with zero mapping quality, suggested a region(s) of very high sequence similarity existed elsewhere in the genome. No paired alignments with greater than expected insert size were present, usually indicative of a deletion.
190x86mm (300 x 300 DPI)
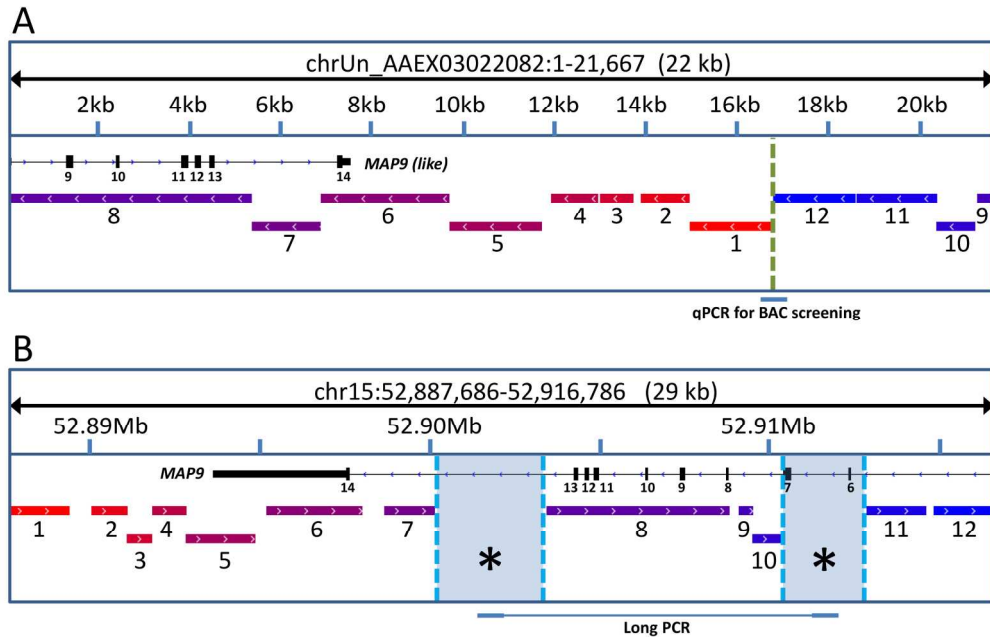
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 2: BLAST results of chrUn_AAEX03022082 to chromosome 15.
Alignment of the chrUn_AAEX03022082 sequence (A) against chromosome 15 (B) using BLAST, revealed two regions which were specific to chromosome 15 (marked with asterisks). Corresponding alignment blocks are number from 1-12 and coloured from red through to blue respectively. An apparent sequence join was identified on chrUn_AAEX03022082 (marked with a green dashed line), suggestive of a structural rearrangement event.

219x139mm (300 x 300 DPI)
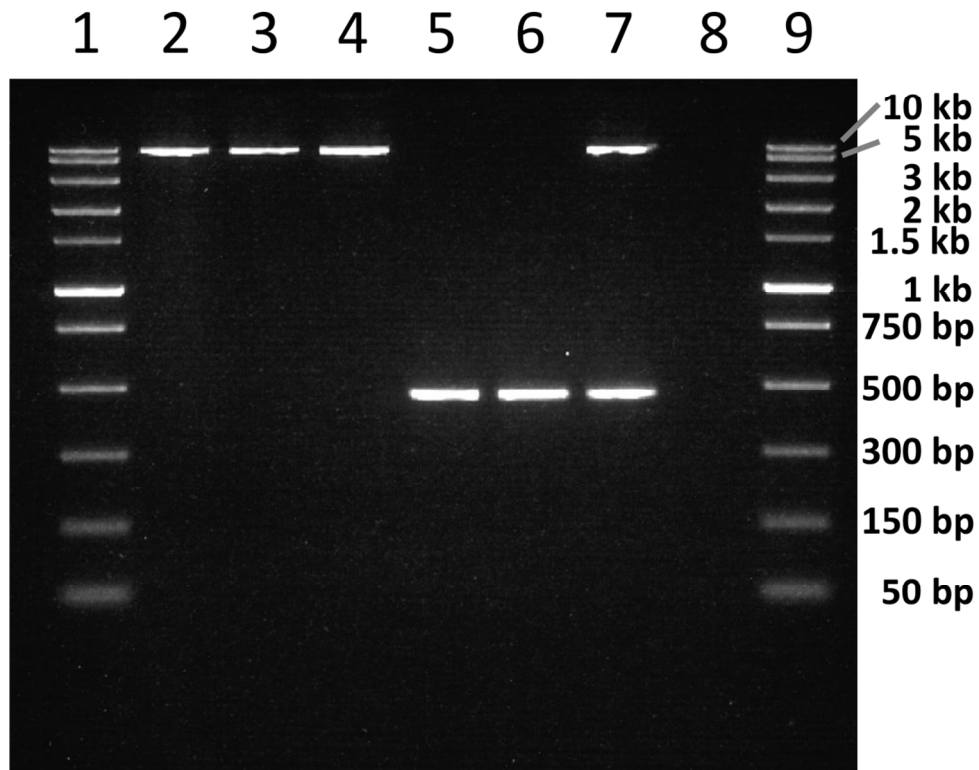
Figure 3: MAP9EORD genotyping PCR
Genotyping of MAP9EORD was carried out by performing one long PCR across the deletion breakpoints and a specific PCR within the deleted region. PCR products were pooled before running on an agarose gel. Lanes 2-4 show a single band representing a PCR product crossing the deletion join (All EORD cases). Lanes 5-6 show a single band representing the PCR within the deleted region. Lane 7 is a heterozygous individual.

118x92mm (300 x 300 DPI)
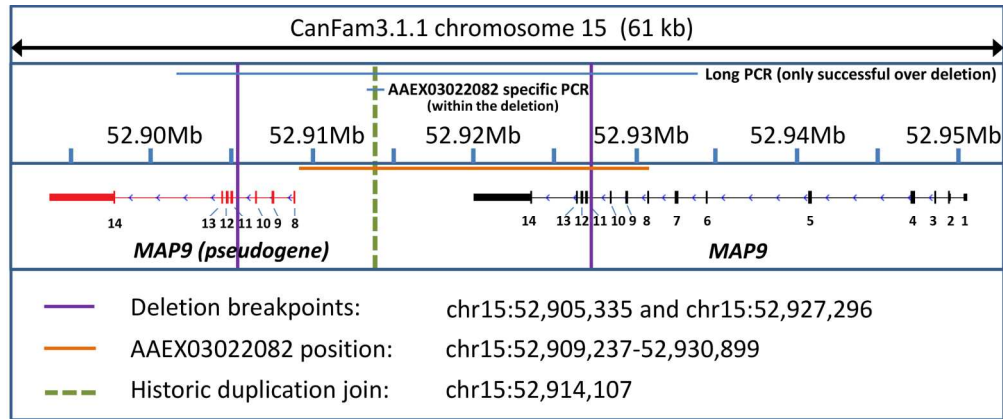
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 4: New assembly arrangement of chromosome 15.
The newly assembled chromosome 15 sequence contains the MAP9 gene and a partial MAP9 pseudogene which we have termed MAP9-like. The arrangement is likely to have occurred through partial tandem duplication of MAP9 caused by homologous recombination.

181x76mm (300 x 300 DPI)
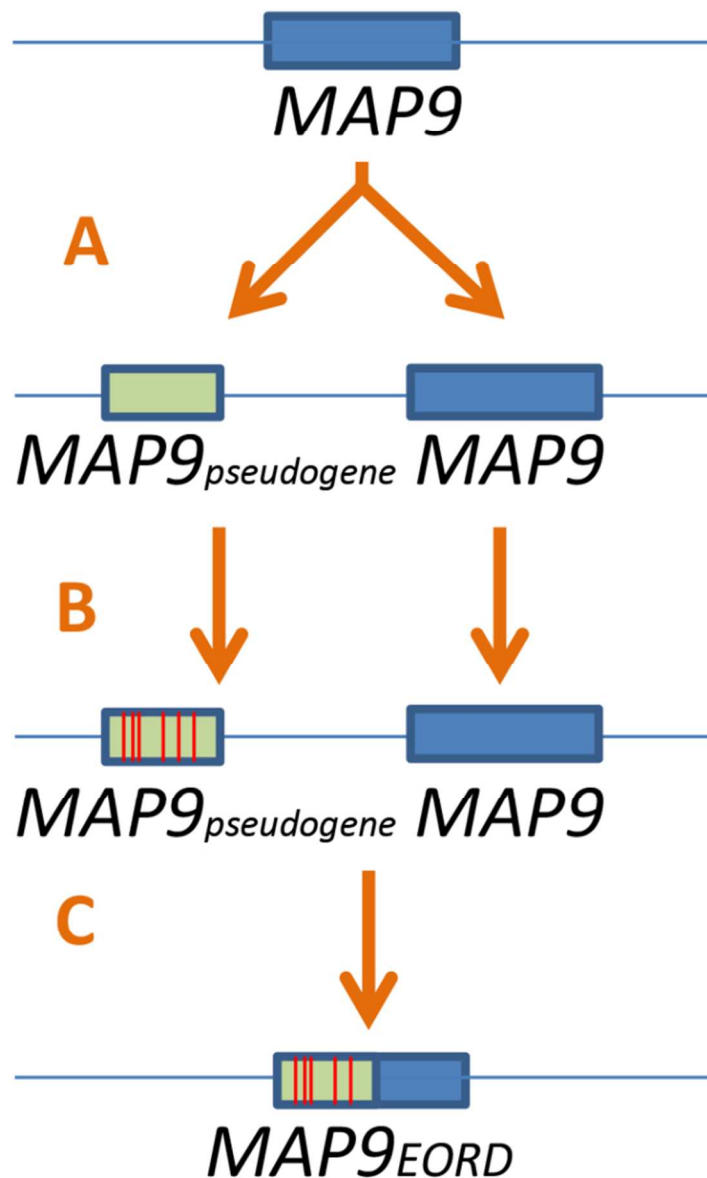
Figure 5: Overview of events leading to MAP9EORD
(A) Homologous recombination event leading to creation of a partial MAP9like pseudogene. This pseudogene is present in all 32 dog breeds investigated for this study, but may be private to the dog, although other closely related canid species have not been investigated. (B) Accumulation of variants in the non-conserved MAP9-like sequence. (C) Homologous recombination event leading to the creation of MAP9EORD. Although this gene is complete and expressed it possesses many variants compared to wildtype MAP9, that change conserved residues of the MAP9 protein.

58x93mm (300 x 300 DPI)

Supplementary File 1

MAP9 cDNA sequencing primers

cDNA_A_F	ACCCGGTCAGAGTTTGAATC
cDNA_A_R	AAGTGGTTGTTGGCTTCAGG
cDNA_B_F	GCCCAGAATTCTTCCAGTCA
cDNA_B_R	GAGGCACTGGACGCTCTATT
cDNA_C_case_F	AATCGAGAAATCCCAGGAAAG
cDNA_C_case_R	TCGGGCAGCTATTTTCTTTG
cDNA_D_case_F	GCCTGGAAGGCTATGAAAGAA
cDNA_D_case_R	CCCCAGATCATCTTCTCCATT
cDNA_C_control_F	GCTGACACAACTGAGTCTTCAAA
cDNA_C_control_R	AATTACATGCCGGCTGTTTC
cDNA_end_F	AAATGGAATGAAAGGAAGGATG
cDNA_end_R	TGTAGTACCCAGATGGCTTCC
cDNA_start_F	CTCAGTCCCGGTCCGTCT
cDNA_start_R	TGGTGACTTTCGGACTCTTTG

Primers for genotyping

Deletion_region_F	AAGTGAATAGCACTTGTGGCTC
Deletion_region_R	TAATTCTGACCGTTTGAGGGTA
Deletion_span_F	TGGGAGCACTTCATTTAACCTCAA
Deletion_span_R	GAAAGCAAATGAATCTGTTATCCACCT

Long PCR for sequencing and denovo assembly to bridge 3' assembly point

Assembly_3prime_F	TGGAGTACCAGAAAGGAAATCAGAAA
Assembly_3prime_R	GAAAGCAAATGAATCTGTTATCCACCT

BAC screening

BAC_probe	/56-FAM/TACACCTCT/ZEN/CCTACTGCCTTACCTTC/3IABkFQ/
BAC_forward	GATGTAGCCGTGACATCTGC
BAC_reverse	GCCGGATAAGTCATTGAAGG

5' MAP9 qPCR

MAP9_5_probe	/56-FAM/AGCTCATCC/Zen/TGGAAAGTAGTCCTTTTGG/3IABkFQ/
MAP9_5_forward	TTTTAGCTTATACAAAGAGTCCGAAAG
MAP9_5_reverse	TTCAGAACTCCTTTGCCTGG

3' MAP9 qPCR

MAP9_3_probe	/56-FAM/TGAACGAAA/ZEN/ACAAAAGAAGCGTCA/3IABkFQ/
MAP9_3_forward	AGGGAGGAAGTGCCTCATTT
MAP9_3_reverse	TGAAAAATGGCTGGAAAGAA

Supplementary File 2
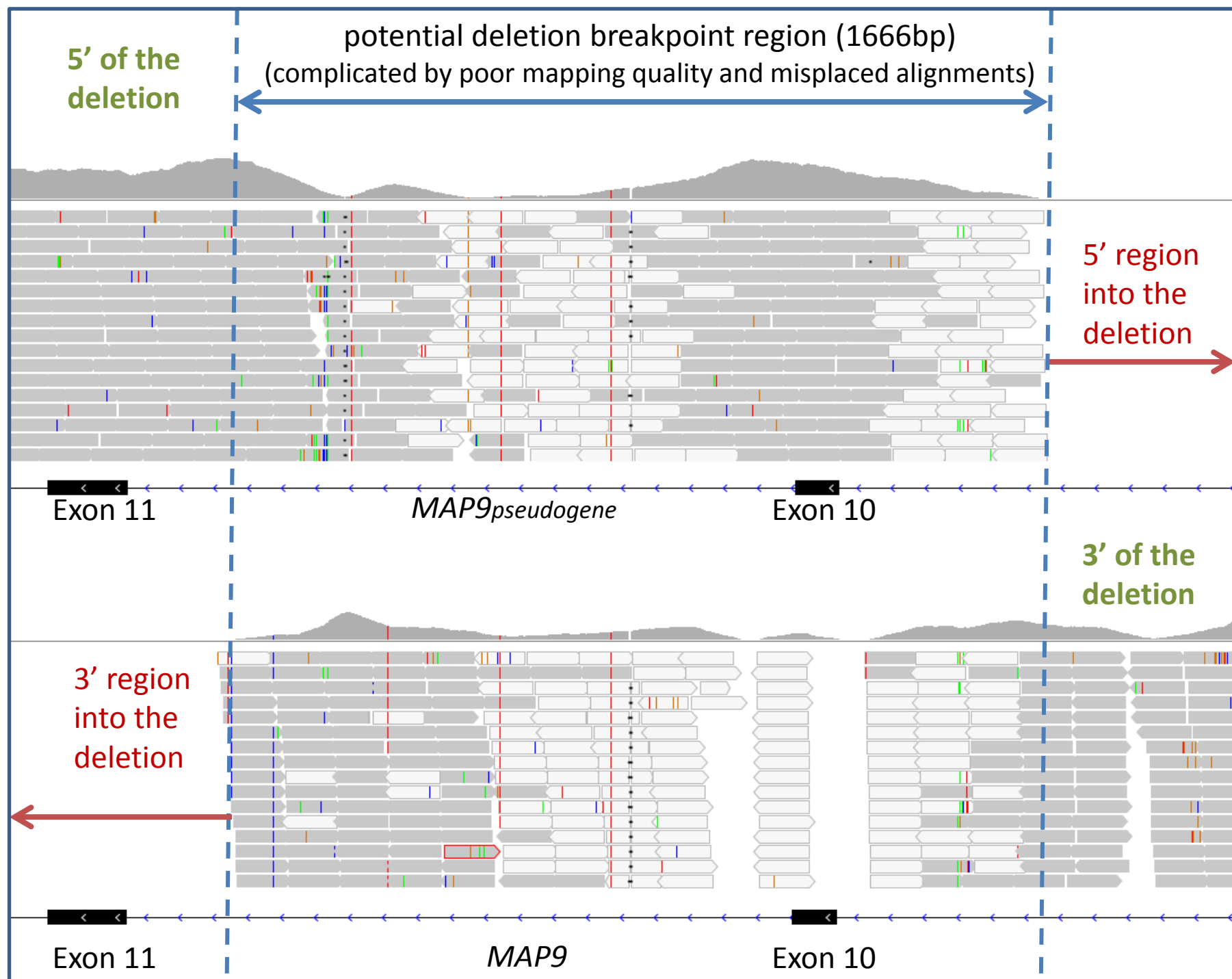
Supplementary File 3

```
CLUSTAL 2.1 multiple sequence alignment
Positions are based on CanFam3.1

N chrUn_AAEX03022082 specific variant
N chr15 specific variant
N polymorphic based on target enrichment data
* Minimal deletion breakpoint region
```

```
EORD                          CAGCTAGACACAAATGCAGGTGTGTGATATGGTGAAATGATGACTGCGGG 50
chr15_52904890-52907293       CAGCTAGACACAAATGCAGGTGTGTGATATGGTGAAATGATGACTGCGGG 50
chrUn_AAEX03022082_1656-4051  CAGCTAGACACAAATGCAGGTGTGTGATATGGTGAAATGATGACTGTGGG 50
                              *********************************************** ***

EORD                          GACTCCAAGCCACTGAGTACCTGCAGCGCCTCTCCCTTTCTGGCAGCATT 100
chr15_52904890-52907293       GACTCCAAGCCACTGAGTACCTGCAGCGCCTCTCCCTTTCTGGCAGCATT 100
chrUn_AAEX03022082_1656-4051  GACTCCAAGCCACTGAGTACCTGAAGCGCCTCTCCCTTTCTGGCAGCATT 100
                              *********************** **************************

EORD                          TTCTTCCTCTGTTTTCTTCTTGTTTTTTTCCTCAAGCCTTTTTCGGGCAG 150
chr15_52904890-52907293       TTCTTCCTCTGTTTTCTTCTTGTTTTTTTCCTCAAGCCTTTTTCGGGCAG 150
chrUn_AAEX03022082_1656-4051  TTCTTCCTCTGTTTTCTTCTTGTTTTTTTCCTCAAGCCTTGTTTGGGCAG 150
                              **************************************** ** ******

EORD                          CTATTTTCTTTGCTTCCTTTTCTTTCATAGCCTTCCAGGCCTCAAAGGAT 200
chr15_52904890-52907293       CTATTTTCTTTGCTTCCTTTTCTTTCATAGCCTTCCAGGCCTCAAAGGAT 200
chrUn_AAEX03022082_1656-4051  CTATTTTCTTTGCTTCCTTTTCTTTCATAGCCTTCCAGGCCTCAAAGGAT 200
                              **************************************************

EORD                          GCTAATGCTTCTTCTCTCTTAGCAGCTCTTTTCTGTTTCAAAAAAAATAAG 250
chr15_52904890-52907293       GCTAATGCTTCTTCTCTCTTAGCAGCTCTTTTCTGTTTCAAAAAAAATAAG 250
chrUn_AAEX03022082_1656-4051  GCTAATGCTTCTTCTCTCTTAGCAGCTCTTTTCTGAT-CAAGAAAATAAG 249
                              *********************************** * *** ********

EORD                          AGACAGTGAGACGTGTGTGTATAAACAGACGGAGACAGGTAGATACGACA 300
chr15_52904890-52907293       AGACAGTGAGACGTGTGTGTATAAACAGACGGAGACAGGTAGATACGACA 300
chrUn_AAEX03022082_1656-4051  AGAC--TGAGACGTGTGTGCATAAACAGACGGAGACAGGTAGATACGACA 297
                              ****  *************** ****************************

EORD                          CACACATATGGATTGTGCTGAATTCCCTTCATAGTCATCATTTCGAGACA 350
chr15_52904890-52907293       CACACATATGGATTGTGCTGAATTCCCTTCATAGTCATCATTTCGAGACA 350
chrUn_AAEX03022082_1656-4051  CACACGTATGGATTGTGCTGAATTCCCTTCATAGTCATCATTTCGAGACA 347
                              ***** ********************************************

EORD                          TTGTTTTGCAGAGCAGGTGACTCATTTAGAATAGTTTAGACAAGCTGATT 400
chr15_52904890-52907293       TTGTTTTGCAGAGCAGGTGACTCATTTAGAATAGTTTAGACAAGCTGATT 400
chrUn_AAEX03022082_1656-4051  TTGTTTCGCAGAGCAGGTGACTTATTTAGAATAGCTTAGACAAGCTGATT 397
                              ****** *************** *********** **************

EORD                          TGGAAAAAAAATGTTACCGGAACTATTCATTGCATTTTAAATATCCTTTT 450
chr15_52904890-52907293       TGGAAAAAAAATGTTACCGGAACTATTCATTGCATTTTAAATATCCTTTT 450
chrUn_AAEX03022082_1656-4051  GGAAAAAAGAATGTCACCAGAACTATTCATTGCATTTAAAATATTCTTTT 447
                              * ***** ***** *** ****************** ****** *****

EORD                          CTAAAACCATTTATCAGATATTGATCTGCTATGAGAAAGACCTCTTTTTT 500
chr15_52904890-52907293       CTAAAACCATTTATCAGATATTGATCTGCTATGAGAAAGACCTCTTTTTT 500
chrUn_AAEX03022082_1656-4051  CTAAAACCATTTATCAGATATTGATCTGCTATGAGAAAGACCTCTTTTTT 497
                              **************************************************

EORD                          GGAAGCGTACTAAAATGTTATCTAAATAACCTATAAAGGCTGAGTGCTCC 550
chr15_52904890-52907293       GGAAGCGTACTAAAATGTTATCTAAATAACCTATAAAGGCTGAGTGCTCC 550
chrUn_AAEX03022082_1656-4051  GGAAGCGTACTAAAATGTTATCTAAATAAACTATAAAGGCTGAGTGCTCC 547
                              ****************************** *****************

EORD                          TGCAAAATGCTCATAATAAAACATTCCATCACAAGAAAAGGATCAGGAAA 600
chr15_52904890-52907293       TGCAAAATGCTCATAATAAAACATTCCATCACAAGAAAAGGATCAGGAAA 600
chrUn_AAEX03022082_1656-4051  TGCAAAATGCTCATAATAAAACATTCCATCACAAGAAAAGGATCAGGAAA 597
                              **************************************************
```

```
1
2
3     EORD                           GTTACAGGTTTATAAAGTAAGTCAGAGTGGGGTAACCCGGCCAAGCCACT 650
4     chr15_52904890-52907293        GTTACAGGTTTATAAAGTAAGTCAGAGTGGGGTAACCCGGCCAAGCCACT 650
5     chrUn_AAEX03022082_1656-4051   GTTACAGGTTTATAAAGTAAGTCAGAGTGGGGTAACCCGGCCAAGCCACT 647
6                                    **************************************************
7     EORD                           TAGGAAGCCAAAATAGATCTTTGCTT-GAACCCGAGTTTTGGTGGACATT 699
8     chr15_52904890-52907293        TAGGAAGCCAAAATAGATCTTTGCTTTGAACCCGAGTTTAGGTGGACATT 700
9     chrUn_AAEX03022082_1656-4051   TAGGAAGCCAAAATAGATCTTTGCTT-GAACCCGAGTTTTGGTGGACATT 696
10                                   **************************  ***********  **********
11    EORD                           AAACCATAAGATTGTGACAACACAAGCAGCACCCAAATATTGAACCTGTA 749
12    chr15_52904890-52907293        AAACCATAAGATTGTGACAACACAAGCAGCACCCAAATATTGAACCTGTA 750
13    chrUn_AAEX03022082_1656-4051   AAACCATAAGATTGTGACAACACAAGCAGCACCCAAATATTGAACCTGTA 746
14                                   **************************************************
15    EORD                           TGTTGGTGAGCATATGGTAGGTATGTTATTACTTGACTTACCTCCATCTA 799
16    chr15_52904890-52907293        TGTTGGTGAGCATATGGTAGGTATGTTATTACTTGACTTACCTCCATCTA 800
17    chrUn_AAEX03022082_1656-4051   TGTTGGTGAGCATACGGTAGGTATGTTATTACTTGACTTACCTCCATCTA 796
18                                   **************  **********************************
19    EORD                           ATCCCATACTATCCTTTGCCTAATAATTCTGTGTAATGTTACACGACATA 849
20    chr15_52904890-52907293        ATCCCATACTATCCTTTGCCTAATAATTCTGTGTAATGTTACACGACATA 850
21    chrUn_AAEX03022082_1656-4051   ATCCCATACTATCCTTTGCCTAATAATTCTGTGTAATGTTACACGACATA 846
22                                   **************************************************
23    EORD                           AAGCCTTACAAAGAGAAGACACTATTTAGGAGTAAAGGTACAAGCTACCT 899
24    chr15_52904890-52907293        AAGCCTTACAAAGAGAAGACACTATTTAGGAGTAAAGGTACAAGCTACCT 900
25    chrUn_AAEX03022082_1656-4051   AAGCCTTACAAAGAGAAGACACTATTTAGGAGTAAAGGTACAAGCTACCT 896
26                                   **************************************************
27    EORD                           CCACGCGGTGTGAAGATCACACGGGAAGGTGTGTGGCACTGAGCACAGCC 949
28    chr15_52904890-52907293        CCACGCGGTGTGAAGATCACACGCAAGGTGTGTGGCACTGAGCACAGCC 950
29    chrUn_AAEX03022082_1656-4051   CCACGCGGTGTGAAGATCACACGGGAAGGTGTGTGGCACTGAGCACAGCC 946
30                                   ***********************  *************************
31    EORD                           TCCTAATGGTGCCGACGTGTTGATGGGGGGTGGGGACCTCCTGGAGGGCT 999
32    chr15_52904890-52907293        TCCTAATGGTGCCGACGTGTTGATGGGGGGTGGGGACCTCCGGGAGGGCT 1000
33    chrUn_AAEX03022082_1656-4051   TCCTAATGGTGCCGACGTGTTGATGGGGGGTGGGGACCTCCGGGAGGGCT 996
34                                   *****************************************  *******
35    EORD                           GCTCCACGCAGCACCATGCATGTGGGAAGTAGCCTTCTTTCCAAGTCCAT 1049
36    chr15_52904890-52907293        GCTCCACGCAGCACCATGCATGTGGGAAGTAGCCTTCTTTCCAAGTCCAT 1050
37    chrUn_AAEX03022082_1656-4051   GCTCCACGCAGCACCATGCATGTGGGAAGTAGCCTTCTTTCCAAGTCCAT 1046
38                                   **************************************************
39    EORD                           TGTAGTGTCGGACAGCACAGGAAAGGCACTTTAGTTAGAAGATGGGAGGG 1099
40    chr15_52904890-52907293        TGTAGTGTCGGACAGCACAGGAAAGGCACTTTAGTTAGAAGATGGGAGGG 1100
41    chrUn_AAEX03022082_1656-4051   TGTAGTGTCGGACAGCACAGGAAAGGCACTTTAGTTAGAAGATGGGAGGG 1096
42                                   **************************************************
43    EORD                           AACTAAGCAGGTTGCAGGCCACCAGGCCTACATGGAGACTGAAGTCGGGT 1149
44    chr15_52904890-52907293        AACTAAGCAGGTTGCAGGCCACCAGGCCTACATGGAGACTGAAGTCGGGT 1150
45    chrUn_AAEX03022082_1656-4051   AACTAAGCAGGTTGCAGGCCACCAGGCCTACATGGAGACTGAAGTCGGGT 1146
46                                   **************************************************
47    EORD                           CAAGAGCTCACATAAGAGCTGGTCTTAAAGACACATATAAAATTCAAGAT 1199
48    chr15_52904890-52907293        CAAGAGCTCACATAAGAGCTGGTCTTAAAGACACATATAAAATTCAAGAT 1200
49    chrUn_AAEX03022082_1656-4051   CAAGAGCTCACATAAGAGCTGGTCTTAAAGACACATATAAAATTCAAGAT 1196
50                                   **************************************************
51    EORD                           GGAAAAGAGGAAAACTGGAGGTGGATGTCTGATGACCCTTCCCATTCTTG 1249
52    chr15_52904890-52907293        GGAAAAGAGGAAAACCGGAGGTGGATGTCTGATGACCCTTCCCATTCTTG 1250
53    chrUn_AAEX03022082_1656-4051   GGAAAAGAGGAAAACCGGAGGTGGATGTCTGATGACCCTTCCCATTCTTG 1246
54                                   ***************  *********************************
55    EORD                           GCTAAGT----CATAAGACTGAAGTCAGTTTTTAATAGTGAAAAGTGAAG 1295
56    chr15_52904890-52907293        GCTAAGTAAGTCATAAGACTGAAGTCAGTTTTTAATAGTGAAAAGTGAAG 1300
57    chrUn_AAEX03022082_1656-4051   GCTAAGTAAGTCATAAGACTGAAGTCAGTTTTTAATAGTGAAAAGTGAAG 1296
58                                   *******    ***************************************
59    EORD                           TAAGAGTTAATAGGTGGTCAGAAGCAATCACATGTTCCTAATCAGTAGTG 1345
60    chr15_52904890-52907293        TAAGAGTTAATAGGTGGTCAGAAGCAATCACATGTTCCTAATCAGTAGTG 1350
      chrUn_AAEX03022082_1656-4051   TAAGAGTTAATAGGTGGTCAGAAGCAATCACATGTTCCTAATCAGTAGTG 1346
                                     **************************************************
```

```
EORD                          TCAGATCTATTCCACAACAACAAAATAGAAATCTCTGTGTATCTGTTTCC 1395
chr15_52904890-52907293       TCAGATCTATTCCACAACAACAAAATAGAAATCTCTGTGTATCTGTTTCC 1400
chrUn_AAEX03022082_1656-4051  TCAGATCTATTCCACAACAACAAAATAGAAATCTCTGTGTATCTGTTTCC 1396
                              *************************************************

EORD                          TTGCGAGAGGATTATATTTATGTCAAACTGAAATGTATGGAGTACCAGAA 1445
chr15_52904890-52907293       TTGCGAGAGGATTATATTTATGTCAAACTGAAATGTATGGAGTACCAGAA 1450
chrUn_AAEX03022082_1656-4051  TTGCGAGAGGATTATATTTATGTCAAACTGAAATGTATGGAGTACCAGAA 1446
                              *************************************************

EORD                          AGGAAATCAGAAATCCATTCTTTTATATACAGGTTATTTTGCCATTTAAA 1495
chr15_52904890-52907293       AGGAAATCAGAAATCCATTCTTTTATATACAGGTTATTTTGCCATTTAAA 1500
chrUn_AAEX03022082_1656-4051  AGGAAATCAGAAATCCATTCTTTTATATACATGTTATTTTGCCATTTAAA 1496
                              ****************************** ******************

EORD                          AACCAGTGGCTTTAGAAATGTTACTGAGAGATATTTTACTGACTTAAAAA 1545
chr15_52904890-52907293       AACCAGTGGCTTTAGAAATGTTACTGAGAGATATTTTACTGACTTAAAAA 1550
chrUn_AAEX03022082_1656-4051  AACCAGTGGCTCTAGAAATGTTACTGAGAGATATTTTACTGACTTAAAAA 1546
                              *********** *************************************

EORD                          TCCAGAATAACTTCCACTTCTATATGTCAGATTACCTGTTCATTTTGGAT 1595
chr15_52904890-52907293       TCCAGAATAACTTCCACTTCTATATGTCAGATTACCTGTTCATTTTGGAT 1600
chrUn_AAEX03022082_1656-4051  TCCAGAATAACTTCCACTTCTATATGTCAGATTACCTGTTCATTTTGGAT 1596
                              *************************************************

EORD                          CCTTAAGTTCTCACTTTCAATCCTTTTAATTCTGTGCATTTCATGTAAAT 1645
chr15_52904890-52907293       CCTTAAGTTCTCACTTTCAATCCTTTTAATTCTGTGCATTTCATGTAAAT 1650
chrUn_AAEX03022082_1656-4051  CCTTAAGTTCTCACTTTCAATCCTTTTAATTCTGTGCATTTCATGTAAAT 1646
                              *************************************************

EORD                          ACACATTTTTCTTTTCTAACCACTCCTGAGAAAATAAGAAAAAAAGAATA 1695
chr15_52904890-52907293       ACACATTTTTCTTTTCTAACCACTCCTGAGAAAATAAGAAAAAAAGAATA 1700
chrUn_AAEX03022082_1656-4051  ACACATTTTTCTTTTCTAACCACTCCTGAGAAAATAAGAAAAAAAGTATA 1696
                              ********************************************** ***

EORD                          GAATGTTCAGTTTCCCAAAAGAATTAGAATTCTTATCCATTTCTGAAATA 1745
chr15_52904890-52907293       GAATGTTCAGTTTCCCAAAAGAATTAGAATTCTTATCCATTTCTGAAATA 1750
chrUn_AAEX03022082_1656-4051  GAATGTTCAGTTTCCCAAAAGAATTAGAATTCTCATCCATTTCTGAAATA 1746
                              ********************************* ***************

EORD                          CTATAAATTTTTAAAACGGCAAACAAAATTTAGAAATTCCATATAAGCCA 1795
chr15_52904890-52907293       CTATAAATTTTTAAAACGGCAAACAAAATTTAGAAATTCCATATAAGCCA 1800
chrUn_AAEX03022082_1656-4051  CTATAAATTTTTAAAACGGCAAACAAAATTTAGAAATTCCATATAAGCCA 1796
                              *************************************************

EORD                          GTATCATGCTTCTTTTTTCTGATAATATTCTAATACGTAAAGACGAGCAT 1845
chr15_52904890-52907293       GTATCATGCTTCTTTTTTCTGATAATATTCTAATACGTAAAGACGAGCAT 1850
chrUn_AAEX03022082_1656-4051  GTATCATGCTTCTTTTTTCTGATAATATTCTAATACGTAAAGACGAGCAT 1846
                              *************************************************

EORD                          GAGTTAAGAAAAATACGTTATCTTTCTTAGTTTTAACAACAGACTTCAAA 1895
chr15_52904890-52907293       GAGTTAAGAAAAATACGTTATCTTTCTTAGTTTTAACAACAGACTTCAAA 1900
chrUn_AAEX03022082_1656-4051  GAGTTAAGAAAAATACGTTATCTTTCTTAGTTTTAACAACAGACTTCAAA 1896
                              *************************************************

EORD                          GATTGAATATAATGGCTACATAAGTTAACATTTTTTTCTCCAAAATTTTT 1945
chr15_52904890-52907293       GATTGAATATAATGGCTACATAAGTTAACATTTTTTTCTCCAAAATTTTT 1950
chrUn_AAEX03022082_1656-4051  GATTGAATATAATGGCTACATAAGTTAACATTTTTTTCTCCAAAATTTTT 1946
                              *************************************************

EORD                          AAGTACAATTTTCTAAAGTGTACTGTTTTGTATAACAAACAAGACCAAAA 1995
chr15_52904890-52907293       AAGTACAATTTTCTAAAGTGTACTGTTTTGTATAACAAACAAGACCAAAA 2000
chrUn_AAEX03022082_1656-4051  AAGTACAATTTTCTAAAGTGTACTGTTTTGTATAACAAACAAGACCAAAA 1996
                              *************************************************

EORD                          AAAGAACCTCTTTTGGAATTGTGAAGACAAATCAAGAAAGTCTTTTTCTC 2045
chr15_52904890-52907293       AAAGAACCTCTTTTGGAATTGTGAAGACAAATCAAGAAAGTCTTTTTCTC 2050
chrUn_AAEX03022082_1656-4051  AAAGAACCTCTTTTGGAATTGTGAAGACAAATCAAGAAAGTCTTTTTCTC 2046
                              *************************************************

EORD                          TGTAAGCTTAATGAAATTACAAGGCCCAAGAAAAAATTAAAACTCTTTTA 2095
chr15_52904890-52907293       TGTAAGCTTAATGAAATTACAAGGCCCAAGAAAAAATTAAAACTCTTTTT 2100
chrUn_AAEX03022082_1656-4051  TGTAAGCTTAATGAAATTACAAGGCCCAAGAAAAAATTAAAACTCTTTTA 2096
                              *************************************************
```

```
EORD                          GTTCTTAACAATATACATGTTAAGAAGAAAAAGAATATAAACATATATAG 2145
chr15_52904890-52907293       GTTCTTAACAATATACATTTTAAGAAGAAAAAGTATAAAAACATATATAG 2150
chrUn_AAEX03022082_1656-4051  GTTCTTAACAATATACATGTTAAGAAGAAAAAGAATATAAACATATATAG 2146
                              ****************** ************** *** ************

EORD                          ACATTAACCACATTTTTGCTTCTACTGTATTTTTATAAAGTGCTGTGGCA 2195
chr15_52904890-52907293       ACATTAACCACATTT--GCTTCTACTGTATTTTTATAAAGTGCTGTGGCA 2198
chrUn_AAEX03022082_1656-4051  ACATTAACCACATTTTTGCTTCTACTGTATTTTTATAAAGTGCTGTGGCA 2196
                              **************   *********************************

EORD                          TATTCTCTTAAACAACTATGATCAGTTAAGCAAAGCTTGGAAGGAAATAA 2245
chr15_52904890-52907293       TATTCTCTTAAACAACTATGATCAGTTAAGCAAAACTTGGAAGGAAATAA 2248
chrUn_AAEX03022082_1656-4051  TATTCTCTTAAACAACTATGATCAGTTAAGCAAAGCTTGGAAGGAAATAA 2246
                              ********************************** ***************

EORD                          TTTAGTAAAAATACCTCTGGGTTGTTCAGATATTATAAAGGAAAAT--TT 2293
chr15_52904890-52907293       TTTAGTAAAAATAACTCTGGGTTGCTCAGATATTATAAAGGAAAATGTTT 2298
chrUn_AAEX03022082_1656-4051  TTTAGTAAAAATACCTCTGGGTTGTTCAGATATTATAAAGGAAAAT--TT 2294
                              ************* ********* *******************   **

EORD                          TTTTTTTTAAGATTTTATTTATT----CATGAGAGACACACAGAGAGGCA 2339
chr15_52904890-52907293       TTTTTTTTAAGATTTTATTTATTTATTCATGAGAGACACACAGAGAGGCA 2348
chrUn_AAEX03022082_1656-4051  TTTTTTTTAAGATTTTATTTATT----CATGAGAGACACACAGAGAGGCA 2340
                              **********************      *********************

EORD                          TAGACACAGGCAGAGGGAGAATCAGGCTCCCTGTGGGGAGCCCGATGTGG 2389
chr15_52904890-52907293       TAGACACAGGCAGAGGGAGAATCAGGCTCCCTGTGGGGAGCCCAATGTGG 2398
chrUn_AAEX03022082_1656-4051  TAGACACAGGCAGAGGGAGAATCAGGCTCCCTGTGGGGAGCCCGATGTGG 2390
                              ******************************************* ******

EORD                          GACTCG 2395
chr15_52904890-52907293       GACTCG 2404
chrUn_AAEX03022082_1656-4051  GACTCG 2396
                              ******
```

Supplementary File 4

MAP9 Residue Number	Amino acid loPRA	Amino acid eoPRA	Amino acid Human (hg19)	PolyPhen prediction	Prodominantly conserved residue	Number of vertebrate species in comparison	vertebrate conservation (%)	Notes
203	R/C*	C	S	Probably damaging	S	63	41	Naturally occuring canine polymorphism
334	I/K*	K	K	Benign	K	57	60	Naturally occuring canine polymorphism
489	Q	R	K	Benign	K	75	80	eoPRA residue in Pika, cape elephant shrew
490	R	K	K	Benign	K	75	96	eoPRA residue is the same as the conserved residue
517	R	Q	K	Possibly damaging	K	70	69	
555	N	K	N	Probably damaging	N	70	84	
576	E	V	E	Benign	E	72	85	eoPRA residue in Tibetan antelope, cow, sheep, domestic goat, manatee.
579	N	S	N	Probably damaging	N	71	73	eoPRA residue in big brown bat
584	E	-	E		E	71	92	Residue also absent in chicken

```
CLUSTAL O(1.2.1) Multiple sequence alignment for MAP9

                Exon  2                          Exon 3
Human   MSDEVFSTTLAYTKSPKVTKRTTFQDELIRAITARSARQRSSEYSDDFDSDEIVSLGDFS
LoPRA   MSDEVFSTILAYTKSPKVTKRTTFQDELIKAITARSARQRSSEYSDDFDSDEIVSLGDFS
EoPRA   MSDEVFSTILAYTKSPKVTKRTTFQDELIKAITARSARQRSSEYSDDFDSDEIVSLGDFS
                                  :
                              Exon 4
Human   DTSADENSVNKKMNDFHISDDEEKNPSKLLFLKTNKSNGNITKDEPVCAIKNEEEMAPDG
LoPRA   DTSVDENSNKKKMNDFHISDDEEKNSPKLSFLKTKKSNSDIMKDEPVFSTKNDEEMAPDG
EoPRA   DTSVDENSNKKKMNDFHISDDEEKNSPKLSFLKTKKSNSDIMKDEPVFSTKNDEEMAPDG
           .    :                    :   .:      :   :
                                              Exon 5
Human   CEDIVVKSFSESQNKDEEFEKDKIKMKPKPRILSIKSTSSAENNSLDTDDHFKPSPRPRS
LoPRA   CGNMVGTPLSESQNNDQEIEKDKIKMKPKPRILPVKSMSSENNSSPEANNHFKPSPRPRS
EoPRA   CGNMVGTPLSESQNNDQEIEKDKIKMKPKPRILPVKSMSSENNSSPEANNHFKPSPRPRS
          ::   .:      : :           :        :  .::::

Human   MLKKKSHMEEKDGLED-KETALSEELELHSAPSSLPTPNGIQLEAEKKAFSENLDPEDSC
LoPRA   MLKKHSHGEEKDGPGEGKTPALREELEARSAPSPLPKLNDGQLEAEKKLASENLDPKDPW
EoPRA   MLKKHSHGEEKDGPGEGKTPALCEELEARSAPSPLPKLNDGQLEAEKKLASENLDPKDPW
           :            :              .         :         : :
                Exon 6                          Exon 7
Human   LTSLASSSLKQILGDSFSPGSEGNASGKDPNEEITENHNSLKSDENKENSFSADHVTTAV
LoPRA   LTSLSSSSLKENLGDSFSPGSGGKASVEDQNEELTENHNSLKSNENEGNSFLIDLVTTPI
EoPRA   LTSLSSSSLKENLGDSFSPGSGGKASVEDQNEELTENHNSLKSNENEGNSFLIDLVTTPI
           :    :            :   :    :      :  :

Human   EKSKESQVTADDLEEEKAKAELIMDDDRTVDPLLSKSQSILISTSATASSKKTIEDRNIK
LoPRA   EKSQESQVITDDLEEEKEKAELIMNDL-TVDPLF-ISQSILISADTTESSKKTVEDRNMK
EoPRA   EKSQESQVITDDLEEEKEKAELIMNDL-TVDPLF-KSQSILISADTTESSKKTVEDRNMK
           :     :           :        :      :.:      :  :
             Exon 8                    Exon 9
Human   NKKSTNNRASSASARLMTSEFLKKSSSKRRTPSTTTSSHYLGTLKVLDQKPSQKQSIEPD
LoPRA   NKKSTNNRASSASGRLMTSEFLKKSSSIRRPPSTTTSSHYLGTLKVLDQKPSQKQNIEPE
EoPRA   NKKSTNNRASSASGRLMTSEFLKKSSSIRRPPSTTTSSHYLGTLKVLDQKPSQKQNIEPE
                      .                                       .  .
                          Exon 10                  Exon 11
Human   RADNIRAAVYQEWLEKKNVYLHEMHRIKRIESENLRIQNEQKKAAKREEALASFEAWKAM
LoPRA   KADSIRAAVYQEWLEKKNVYLHEMHRIKRIESENLRIQNEQKRAAKREEALASFEAWKAM
EoPRA   KADSIRAAVYQEWLEKKNVYLHEMHRIKRIESENLRIQNEQKRAAKREEALASFEAWKAM
          :  .                                    :
                                              Exon 12
Human   KEKEAKKIAAKKRLEEKNKKKTEEENAARKGEALQAFEKWKEKKMEYLKEKNRKEREYER
LoPRA   KEKEAKKIAAQRRLEEKNKKKTEEENAARKGEALQAFERWKEKKMEYLREKNKKEREYER
EoPRA   KEKEAKKIAARKRLEEKNKKKTEEENAARKGEALQAFECWKEKKMEYLREKNKKEREYER
                   ::                             :      :   :
                                       Exon 13
Human   AKKQKEEETVAEKKKDNLTAVEKWNEKKEAFFKQKEKEKINEKRKEELKRAEKKDKDKQA
LoPRA   AKKQKEEETIAEKRKDNLTAIEKWNERKDAFFKEKEKEKINEKRREELKRAEKKDKDKQA
EoPRA   AKKQKEEETIAEKRKDKLTAIEKWNERKDAFFKEKEKEKISEKRRE-LKRAEKKDKDKQA
                   :    :   :          :          .  :    .
                              Exon 14
Human   INEYEKWLENKEKQERIERKQKKRHSFLESEALPPWSPPSRTVFAKVF
LoPRA   IDEYEKWLEKKEKQERIERKQKKRMSFLNEALPPWSPPSRTVFSRVF
EoPRA   IDEYEKWLEKKERQERIERKQKKRHSFLENEALPPWSPPSRTVFSRVF
           :        :  :                .          ::
```
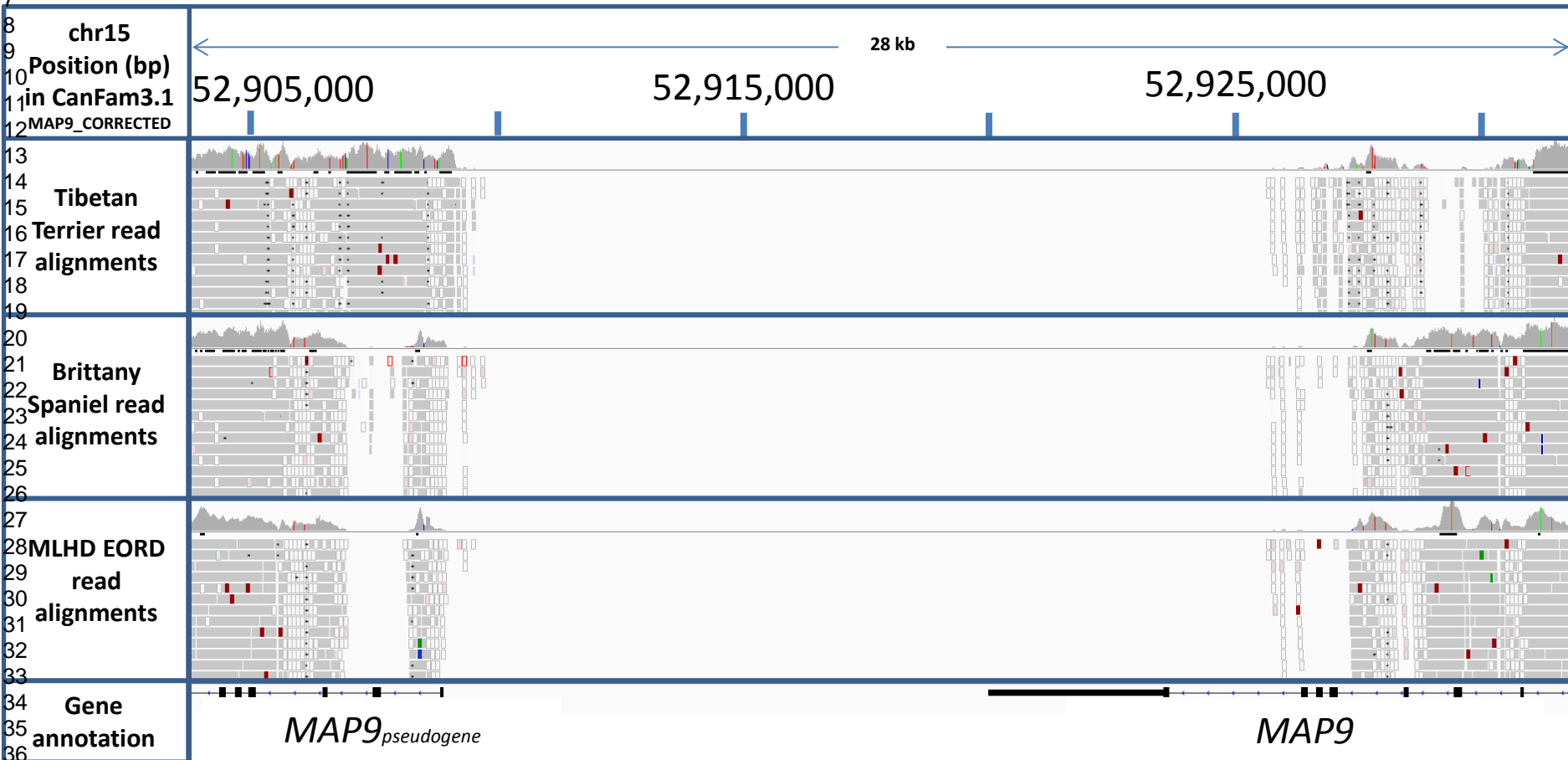
Supplementary File 5

MAP9 deletion genotyping results

Breed	ID	Category of onset	Age of diagnosis	Top SNP Genotype	*MAP9* Genotype
MLHD	7921	E	0.3	G/G	wt/wt
MLHD	7912	E	0.5	T/T	N/A
MLHD	MLD27	E	1.0	T/T	N/A
MLHD	MLD37	E	1.0	T/T	del/del
MLHD	MLD28	E	1.5	T/T	del/del
MLHD	MLD29	E	1.3	T/T	del/del
MLHD	MLD30	E	1.8	T/T	del/del
MLHD	MLD31	E	1.8	T/T	del/del
MLHD	7916	E	1.8	T/T	del/del
MLHD	7920	E	1.8	T/T	del/del
MLHD	5898	E	1.9	T/T	del/del
MLHD	MLD34	E	<2.0	T/T	del/del
MLHD	MLD66	E	2.0	T/T	del/del
MLHD	6088	E	2.2	T/T	del/del
MLHD	17795	E	2.6	G/G	wt/wt
MLHD	MLD2	E	2.7	T/T	del/del
MLHD	MLD32	E	3.0	T/T	del/del
MLHD	MLD70	E	3.0	T/T	del/del
MLHD	7918	E	3.0	T/G	wt/del
MLHD	MLD64	E	<3.4	T/T	del/del
MLHD	MLD25	E	4.0	T/T	del/del
MLHD	7910	E	4.0	T/T	del/del
MLHD	MLD4	E	2.4-5.2	T/T	del/del
MLHD	MLD21	E	5.0	T/T	del/del
MLHD	MLD23	E	5.0	T/T	del/del
MLHD	MLD9	L	5.2-6.6	T/T	del/del
MLHD	MLD291	L	5.6	G/G	wt/wt
MLHD	MLD10	L	Normal @ 6.1	T/G	wt/del
MLHD	11044	L	6.3	G/G	wt/wt
MLHD	MLD39	L	7.0	T/G	wt/wt
MLHD	11056	L	7.7	T/G	wt/wt
MLHD	MLD3	L	7.8-9.3	T/G	wt/del
MLHD	MLD61	L	8.0	T/G	wt/del
MLHD	MLD65	L	SARD @ 8.0	T/G	wt/del
MLHD	MLD132	L	Normal @ 8.5	T/G	wt/del
MLHD	11057	L	8.8	T/G	wt/wt
MLHD	MLD6	L	Normal @ 8.9	G/G	wt/wt
MLHD	MLD24	L	9.0	T/G	wt/wt
MLHD	MLD1	L	9.2-10.6	T/G	wt/del
MLHD	MLD7	L	9.4	T/G	wt/del
MLHD	MLD5	L	>10.8	T/T	wt/del
MLHD	MLD153	L	Normal @ 10.8	G/G	wt/wt
MWHD	11063	L	5.4	G/G	wt/wt
MWHD	11058	L	6.4	G/G	wt/wt
MWHD	11062	L	7.7	T/G	wt/del
CCR	22329	E	2.0	G/G	wt/wt
ESS	6860	E (GPRA)	2.7	T/G	wt/wt
ESS	6862	E (GPRA)	2.7	T/G	wt/wt

Key	
E	Early onset PRA
L	Late onset PRA
MLHD	Miniature Long Haired Dachshund
MWHD	Miniature Wire Haired Dachshund
CCR	Curly Coat Retriever
ESS	English Springer Spaniel

1
2
3
Supplementary File 6
4
5
6
7
8

chr15 Position (bp) in CanFam3.1	28 kb
	52,905,000 52,915,000 52,925,000

MAP9_CORRECTED

Tibetan Terrier read alignments

Brittany Spaniel read alignments

MLHD EORD read alignments

Gene annotation

MAP9~pseudogene~

MAP9