

## Canine *RD3* mutation establishes rod-cone dysplasia type 2 (*rcd2*) as ortholog of human and murine *rd3*

Anna V. Kukekova · Orly Goldstein · Jennifer L. Johnson · Malcolm A. Richardson · Susan E. Pearce-Kelling · Anand Swaroop · James S. Friedman · Gustavo D. Aguirre · Gregory M. Acland

Received: 7 August 2008 / Accepted: 19 November 2008 / Published online: 9 January 2009  
© Springer Science+Business Media, LLC 2008

**Abstract** Rod-cone dysplasia type 2 (*rcd2*) is an autosomal recessive disorder that segregates in collie dogs. Linkage disequilibrium and meiotic linkage mapping were combined to take advantage of population structure within this breed and to fine map *rcd2* to a 230-kb candidate region that included the gene *Clorf36* responsible for human and murine *rd3*, and within which all affected dogs were homozygous for one haplotype. In one of three

identified canine retinal *RD3* splice variants, an insertion was found that cosegregates with *rcd2* and is predicted to alter the last 61 codons of the normal open reading frame and further extend the open reading frame. Thus, combined meiotic linkage and LD mapping within a single canine breed can yield critical reduction of the disease interval when appropriate advantage is taken of within-breed population structure. This should permit a similar approach to tackle other hereditary traits that segregate in single closed populations.

A. V. Kukekova and O. Goldstein contributed equally to this work. Nucleotide sequence data reported here are available in the GenBank database under accession numbers EU687743, EU687744, and EU687745.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00335-008-9163-4) contains supplementary material, which is available to authorized users.

A. V. Kukekova · O. Goldstein · J. L. Johnson · M. A. Richardson · S. E. Pearce-Kelling · G. M. Acland (✉)  
James A. Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University, Hungerford Hill Road, Ithaca, New York 14853, USA  
e-mail: gma2@cornell.edu

M. A. Richardson  
Department of Microbiology and Immunology, Dartmouth Medical School, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire 03766, USA

A. Swaroop  
W. K. Kellogg Eye Center, University of Michigan, Ann Arbor, Michigan 48105, USA

A. Swaroop · J. S. Friedman  
National Eye Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

G. D. Aguirre  
School of Veterinary Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

### Introduction

The development of modern dog breeds has created a population structure that is largely separated into relatively closed subpopulations termed breeds. The traits that define each such breed include both those deliberately bred for and undesirable traits concentrated into particular breeds by descent from a small founder pool. As information about the canine genome became increasingly available, it quickly became apparent that, in general, linkage disequilibrium (LD) within breeds can spread to several megabases, but across breeds LD extends over only tens of kilobases (Lindblad-Toh et al. 2005). This has suggested that the dog is particularly advantageous for gene-mapping studies because the large LD within a breed would require a correspondingly smaller number of markers for association mapping studies in dogs compared to humans (Lindblad-Toh et al. 2005; Sutter et al. 2004). Furthermore, when a trait segregates in multiple breeds, fine mapping across these breeds can very efficiently narrow the candidate region by taking advantage of the reduced LD interval common to all such breeds (Karlsson et al. 2007). Several

identical-by-descent (IBD) mutations have been fine mapped in this manner (Candille et al. 2007; Goldstein et al. 2006; Neff et al. 2004; Parker et al. 2007; Sutter et al. 2007). In previous studies we employed this strategy to fine map two canine hereditary retinal disorders, collie eye anomaly (Parker et al. 2007) and progressive rod-cone degeneration (Goldstein et al. 2006). Both of these diseases segregated in multiple breeds, and fine mapping between such breeds rapidly reduced the initially identified candidate region to yield the subsequently identified causative mutations (Parker et al. 2007; Zangerl et al. 2006). Retrospective analysis of these studies, however, suggested that in both cases a careful selection of individuals based on population structure would have resulted in the same progress, even if restricted to a single breed per disease. The significance of this observation has been strengthened by independent reports of stratified linkage disequilibrium within single breeds (Bjornerfeldt et al. 2008; Quignon et al. 2007). Successful implementation of this strategy would facilitate mapping and fine mapping of traits that are known to segregate only in single specific breeds.

Rod-cone dysplasia type 2 (*rcd2*) is one of the many canine hereditary retinal degenerations that are collectively named progressive retinal atrophy (PRA) (Aguirre and Acland 2006). Some forms of PRA are common to multiple dog breeds, while others are recognized in just a single breed (Acland et al. 1998, 1999; Goldstein et al. 2006; Kijas et al. 2002; Mellersh et al. 2006; Zangerl et al. 2006; Zeiss et al. 2000). As far as is known, *rcd2* segregates exclusively in rough and smooth collies (Wolf et al. 1978). The disease is inherited as a simple autosomal recessive trait and has been previously characterized electrophysiologically, morphologically, and biochemically (Acland et al. 1989; Chader et al. 1985; Santos-Anderson et al. 1980; Woodford et al. 1982). Night blindness is the earliest clinical sign detectable in 6-week-old affected dogs. Retinal dysfunction can be detected by electroretinography (ERG) as early as 16 days of age. At 6 weeks of age, when the photoreceptors of normal dogs are fully developed, only a few underdeveloped outer segments are visible in *rcd2* dogs; by 2–2.5 months age, the outer segments completely disappear in the affected retina. Both rods and cones in the affected retina fail to develop normal outer segments (Santos-Anderson et al. 1980). Both types of photoreceptors subsequently degenerate, cones more slowly than rods. Ophthalmoscopic abnormalities can be detected at 3.5–4 months of age, including tapetal hyperreflectivity, retinal vascular attenuation, and optic nerve pallor. By 6–8 months of age, *rcd2* dogs become functionally blind. The *rcd2* locus has been mapped previously to an approximately 4-Mb region on CFA7, an interval homologous to human 1q32 (Kukekova et al. 2006).

In the present study a combination of meiotic linkage and linkage disequilibrium mapping, together with critical

pedigree analysis, was employed to fine map and reduce this interval. Meiotic mapping used both additional, newly developed, informative pedigrees and additional microsatellite markers located within the previously identified zero-recombination region. In parallel, a dense SNP map of the region was simultaneously developed to permit haplotype analysis and linkage disequilibrium mapping. In combination, these two approaches reduced the candidate region to an approximately 230-kb interval on CFA7 that contained only three known genes (*TRAF5*, *C1orf36*, and *SLC30A1*). All three genes are retinally expressed and were thus potential positional candidates for *rcd2*. Comparative genomic analysis disclosed that the reduced canine *rcd2* interval overlapped with the murine retinal degeneration 3 (*rd3*) locus on Mmu1 (Danciger et al. 1999). Simultaneously, and independently, the *C1orf36* gene was identified as harboring the causative mutations for both murine *rd3* and the homologous human disease (Friedman et al. 2006). This gene has now been renamed *RD3*. Sequence of the canine *RD3* coding region was retrieved from retinal cDNA. Unlike mouse and human *RD3*, which each have a single known transcript (Friedman et al. 2006; Lavorgna et al. 2003), three canine retinal *RD3* splice variants were detected. A sequence alteration identified in one of these variants between normal and *rcd2*-affected individuals is proposed as the cause of canine *rcd2*.

## Methods

To fine map and reduce the *rcd2* candidate region, a two-pronged approach was adopted. First, for meiotic linkage purposes the previously mapped families were significantly expanded (see below) to provide an additional 140 informative progeny. Both the previously mapped and newly developed pedigrees were then genotyped for a much denser set of microsatellites identified from the canine genome assembly as located within the region. Second, and simultaneously, a dense SNP map of the candidate region was developed to permit haplotype analysis and linkage disequilibrium mapping of the candidate region.

### Pedigrees and DNA samples

Blood and/or tissue samples were obtained from dogs in *rcd2*-informative collie-derived mixed-breed pedigrees developed at the Retinal Disease Studies Facility in Kennett Square, PA. Additional blood samples for DNA analysis were obtained from privately owned dogs representative of the United States' pedigreed collie population. These pedigreed collies included both rough and smooth varieties of the breed. It should be noted that internationally these varieties are classified as separate breeds and are

not bred to each other, but they form a single breed registry in the U.S. Extended pedigrees of *rcd2*-affected collies were obtained from historical records provided by collie breeders and analyzed by inspection to detect patterns of common ancestry. Further blood samples for DNA analysis were obtained for population studies from privately owned dogs representing 19 other breeds. Samples used for meiotic linkage analysis represented 14 three-generation pedigrees (126 individuals in the informative generation) previously used for an *rcd2* genome-wide scan (Kukekova et al. 2006), and an additional 22 backcross and intercross pedigrees, including 140 informative progeny. LD mapping was performed using samples from seven *rcd2*-affected and three *rcd2*-carrier (obligate-heterozygous) collie dogs. DNA from canine tissue samples (blood, spleen) was extracted using either Qiagen Mini and Maxi Blood kits (Qiagen, Valencia, CA) or phenol-chloroform extraction methods (Gilbert and Vance 1994).

#### Microsatellite markers and linkage analysis

Three microsatellite markers (FH2226, FH3972, and VIASD10) that defined the previously identified *rcd2* interval (Kukekova et al. 2006) were chosen, plus a fourth (CPH20) from the corresponding interval of the canine integrated genome map (Guyon et al. 2003). Potential additional microsatellites from this interval were identified from CanFam2 (May 2005 assembly of the 7.6x canine genome sequence) using the RepeatMasker track of the UCSC browser (<http://genome.ucsc.edu/cgi-bin/hgGateway?org=Dog&db=canFam2>), and primer pairs for each were designed using Primer3. Such potential markers were tested for informativeness on DNA representing four dogs from the parental generation of *rcd2*-informative pedigrees using standard PCR conditions: an initial 2-min denaturation at 96°C; then 30 cycles of 96°C (20 sec), 58°C (20 sec), and 72°C (20 sec); and a final extension step at 72°C for 5 min. PCR products were resolved on 10% native polyacrylamide gel and visualized by ethidium bromide staining. Five such markers that proved polymorphic were selected for genotyping (SYT14F3/R3, SERTAD4F1/R1, KCNH1F3/R3, *rcd2ms3F1/R1*, *rcd2ms6F1/R1*; see Table 1).

For fine mapping and linkage analysis, pedigrees were genotyped with markers FH2226, CPH20, SYT14F3/R3, KCNH1F3/R3, FH3972, and VIASD10 using fluorescently labeled primers. PCRs were performed in a 15- $\mu$ l mixture containing 1 $\times$  Taq polymerase buffer (Invitrogen, Carlsbad, CA), 1.5 mM MgCl<sub>2</sub>, 0.2 mM dNTP, 0.3 pmol of each primer, 1.5 ng canine DNA, and 0.5 units of Taq polymerase (Invitrogen). Amplification conditions were as described above for amplification with unlabeled primers except that the final extension time was for 1 h. PCR products were combined into a multiplex set and analyzed

**Table 1** Newly designed microsatellite markers and their locations on canine chromosome 7 (CFA7) in the CanFam2 assembly of the dog genome sequence

Primer	Primer sequence	Amplicon location
SYT14F3	ctgcctgtgtctctcctctctc	11,599,574–11,599,748
SYT14R3	aatgtggagccccatttcaag	
SERTAD4F1	accctttctctccatgctagtgc	11,816,877–11,817,110
SERTAD4R1	gcctttctgaacacactagtgacc	
KCNH1F3	ggtcttttagcttccaactgcaga	12,568,635–12,568,820
KCNH1R3	ctaactgaatcagggtgacaga	
<i>rcd2ms6F1</i>	tccagaattcctctccatctcag	12,642,574–12,642,844
<i>rcd2ms6R1</i>	tccacagggtttgtctcaaaaga	
<i>rcd2ms3F1</i>	caaggatgggtccctaaacaga	12,841,258–12,841,496
<i>rcd2ms3R1</i>	cagagggtctgacctatggcaagt	

on an ABI3730 capillary-based genetic analyzer (Applied Biosystems, Foster City, CA). PCR products were sized relative to an internal size standard (GeneScan 500 LIZ) using Genemapper 3.5 software package (Applied Biosystems). Quality parameters were established using Genemapper 3.5 and genotypes were double scored by independent investigators.

#### SNP markers

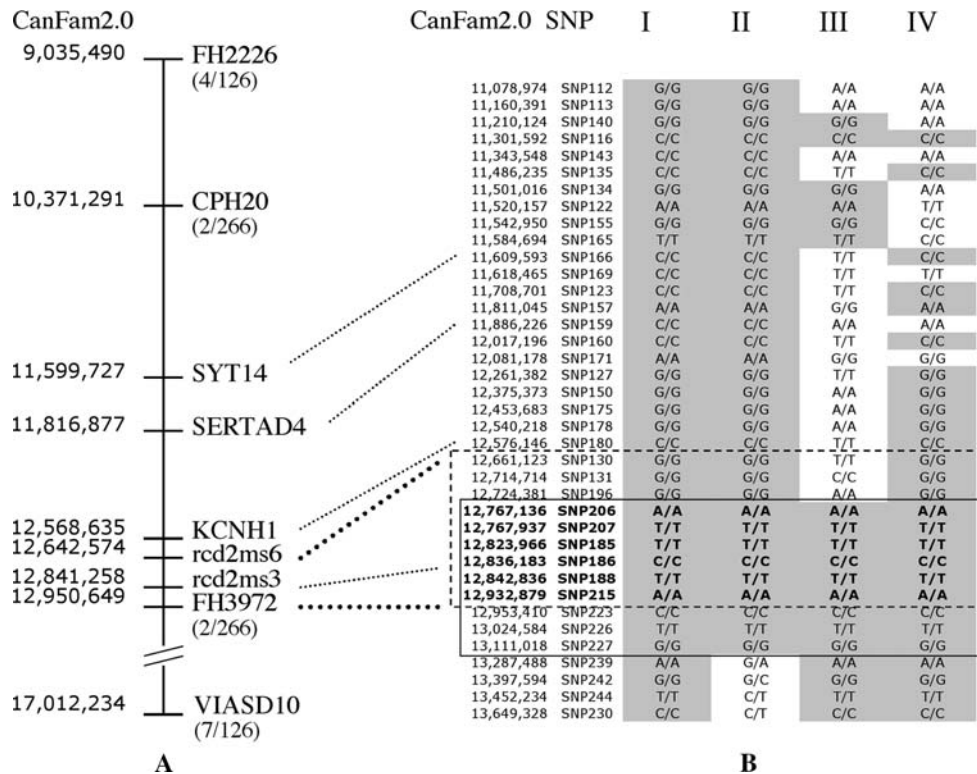
SNP markers for the *rcd2* interval were selected from the canine SNP database (<http://www.broad.mit.edu/mammals/dog/snp/>). Primers for amplification were designed using Primer3 ([http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)). Identified SNPs were amplified on DNA from three to seven *rcd2*-affected and one to three carrier dogs using the same PCR conditions described above for PCR with unlabeled primers. Robustly amplified PCR products were sequenced and analyzed using Sequencher software (GeneCodes Corp., Ann Arbor, MI). Informative SNPs were used for haplotype analysis (Fig. 1b, Supplementary Table 1).

#### RNA

Total RNA was extracted from retinas of normal and *rcd2*-affected dogs using TRIZOL reagent (Ambion, Austin, TX). Time points included 1.4–1.9 weeks, 3.3–4.3 weeks, 6.4–7.9 weeks, and 9.7–10.3 weeks with a normal and *rcd2*-affected dog for each time point.

#### RT-PCR

To evaluate the *rcd2* gene, primers for RT-PCR were designed from *RD3* sequence retrieved from the 7.6 $\times$  assembly of the canine genome (CanFam2) to cover the



**Fig. 1** High-resolution meiotic linkage (a) and linkage disequilibrium (b) maps of the *rcd2* interval on canine chromosome 7. (a) The map order of nine microsatellites is indicated with their sequence locations in the CanFam2 assembly of the canine genome. For microsatellites FH2226, CPH20, FH3972, and VIASD10, the number of recombination events observed between each marker and *rcd2* per total number of informative individuals analyzed (126) is listed in parentheses below each marker's name. SYT14, SERTAD4, KCNH1, and *rcd2ms6* all cosegregated with each other, sharing one observed recombination with *rcd2*. Marker *rcd2ms3* cosegregated with *rcd2* with no observed recombinations. (b) CanFam2 locations and

identifying names are listed for 38 SNPs identified in the *rcd2* candidate region, as are genotypes for these SNPs from four *rcd2*-affected pedigreed collie dogs (I–IV). Shaded regions indicate where all four dogs share identical SNP genotypes. The region encompassed by the dashed box indicates the final *rcd2* zero-recombination interval defined by the locations of markers *rcd2ms6* and FH3972. The region boxed with a solid outline indicates the minimal region of absolute common LD, defined by the locations of SNP196 and SNP239. The final *rcd2* candidate region, representing the overlap of the zero-recombination and absolute common LD intervals, is indicated by SNP genotypes in bold

complete coding sequence (Table 2). Since the predicted exon 4 of this gene was missing from CanFam2 and was not found in canine sequences from any other database, primers within exon 4 were designed from human *RD3* sequence in regions highly conserved among human, mouse, chicken, and opossum genomes. Once the canine sequence was retrieved from RT-PCR reactions, canine primers were designed to further improve the efficiency of reactions (Table 2A). cDNA was prepared using Thermoscript RT (Invitrogen) following the manufacturer's protocol using an extension temperature of 60°C. PCR reactions were run using GoTaq Green Master Mix (Promega, Madison, WI) following the manufacturer's protocol with 5% DMSO and annealing temperature in the range of 62–66°C.

To identify the 5' and 3' ends of the transcript, 5' RACE and 3' RACE were undertaken on 10.4-week-old normal canine retinal RNA using the RACE kit BD Advantage following the manufacturer's protocol (Clontech Laboratories, Mountain View, CA). 5' RACE-PCR with primer

36F9 and 3' RACE-PCR with primer 36R1A were performed. PCR products were cloned into TOPO vector pCR2.1 (Invitrogen). Two 3' RACE and a 5' RACE sequence were aligned using Sequencher to identify an initial consensus canine *RD3* cDNA sequence.

*TRAF5* retinal cDNA of one normal and one *rcd2*-affected dog was amplified using RT-PCR with primers TR1F/TR1R, TR4F/TR3R, TR7F/TR10R, TR4F/TR7R, and TR6F/TR9R (Supplementary Table 2; NCBI accession No. EU687744).

Partial canine *SLC30A1* retinal cDNA was amplified using RT-PCR with primers SL1F/SL3F/SL2R, and SL6F/SL6R. Products corresponding to coding regions of the *SLC30A1* gene were also amplified from canine genomic DNA using primers SL1F/SL1R, SL4F/SL3R, and SL5F/SL5R (Supplementary Table 2). All products were amplified from retinal cDNA or genomic DNA from at least one normal and one *rcd2*-affected dog and sequenced (NCBI accession No. EU687743).

**Table 2** *RD3* primers used for amplification of canine cDNA (A), for *rcd2* mutation screening (B), and to develop hybridization probes (C)

Primer	Sequence (5' to 3')	Source	Location	CanFam2.0
<b>A. Primers used to amplify canine <i>RD3</i> cDNA</b>				
36R1A	gtggatcttggcacagacgtcctct	Canine	Canine exon 3	12,835,497–521
36F9	cgcagtaacgcgctcaggaaggtct	Canine	Canine exon 3	12,835,601–625
36F18	ccaggctggatccctccggctca	Canine	Canine exon 1	12,847,140–161
36R16	agtcatctgaaggcactccagacct	Canine	Canine exon 6	12,820,212–236
36JR2	gcagcgactacgcccctcaggaacaat	Canine	3' UTR v2	12,832,366–391
36e2R2	cgctccacgtcctcggagatggt	Human	Human exon2	
36F2	gtgctggagacgctcatgatgga	Canine	Canine exon 3	12,835,669–691
36R1	gatcttggcacagacgtcctcta	Canine	Canine exon 3	12,835,500–522
36e2F3og	cggtcttccgctcggctgct	Canine	Canine exon 4	Missing in CanFam2.0
36e2ogR16	ggcgcccgaattcgggcatg	Human	Human exon 2	
36e2ogR13	cggttggggcggccggaatc	Human	Human exon 2	
36BACR2	ggaaggcccgtccctcccgtg	Canine	3' UTR v3	Missing in CanFam2.0
36JR3	ctgaatctccagagcccgacacc	Canine	Canine exon 4	Missing in CanFam2.0
36e2F7og	ctgcccctcggccggccacatc	Canine	Canine exon 4	Missing in CanFam2.0
36e2F30	ccccggcctggagcctgcg	Canine	Canine exon 4	Missing in CanFam2.0
36e2R30	gcaggctccaggtgcgagc	Canine	Canine exon 4	Missing in CanFam2.0
36e2R31	cggaaactcggcaggtccaggtg	Canine	Canine exon 4	Missing in CanFam2.0
36e2R32	cctgcggcggccggaactcg	Canine	Canine exon 4	Missing in CanFam2.0
<b>B. Primers used to screen the <i>rcd2</i> mutation</b>				
RD3NH3F1 <sup>a</sup>	tgagtgtgagtttgaggtgatgagg	Canine	Canine exon 4	Missing in CanFam2.0
RD3NH3R1 <sup>a</sup>	cacacacaaaaccacaaaatcc	Canine	Canine exon 4	Missing in CanFam2.0
<b>C. Primers used to produce PCR probes for BAC DNA hybridization</b>				
RD3F2	gtggtgtccatcctctgctggt	Canine	Canine exon 5	12,832,565–586
RD3R2	ccctacaggctcctcgtcctt	Canine	Intronic sequence	12,832,000–021
RD3F4	tgcaagagcccatagaacacca	Canine	Intronic sequence	12,835,201–222
RD3R4	gtaaagggatgcatccaccg	Canine	Intronic sequence	12,834,685–706
36e2F1c	gttccggcagctgctggccga	Canine	Canine exon 4	Missing in CanFam2.0
36e2ogR16	ggcgcccgaattcgggcatg	Human	Human exon 2	

<sup>a</sup> Primers RD3NH3F1 and RD3NH3R1 were designed for amplification of deaminated DNA

### Northern analysis

Northern analysis was performed as described previously (Goldstein et al. 2006). Eleven retinal RNA probes from time points ranging from 1.4 to 10.3 weeks old were analyzed. An *RD3* probe was produced by amplification of cDNA with primers 36F18/36R1. The resulting probe comprises a *RD3* cDNA fragment from exon 1 through exon 3 (primer location is listed in Table 2). The 737-bp PCR product was gel-purified, cloned (TOPO TA cloning kit, Invitrogen) and used for blot hybridization. Hybridization was performed using Ultrahyb solution (Ambion) following the manufacturer's protocol. The blot was exposed to X-ray film at  $-70^{\circ}\text{C}$  for 72 h using two intensifying screens. Loading control was achieved by hybridizing a canine  $\beta$ -actin probe to the membrane under the same conditions and exposure as to X-ray film for 5 h.

Northern images were scanned on a Fuji Bio-Imaging Analyzer and scored by MacBAS software.

### Cloning and sequencing of the gap region

To obtain the sequence representing the genomic fragment of canine *RD3* that is absent from the CanFam2 assembly, two canine BAC clones CH82-166H15 and CH82-417I05 (Boxer BAC library, <http://bacpac.chori.org/library.php?id=253>) were selected from the UCSC database (<http://genome.ucsc.edu/>). BAC DNA was digested with selected restriction enzymes and hybridized with canine *RD3* probes #1 (RD3F2/R2) and #2 (RD3F4/R4) located before and after the gap in the CanFam2.0, respectively, and probe #3 (36e2F1c/36e2ogR16) corresponding to a part of the canine *RD3* exon 4. Primers used for probe amplification are listed in Table 2C. A *Pst*I fragment, about 1.5 kb and

identified from BAC clone CH82-166H15 and tested positive for probes 1 and 3, and a *Pst*I fragment, about 4.5 kb and identified from BAC clone CH82-417I05 that was positive for probes 2 and 3, were both cloned into vector pUC19 at the corresponding restriction sites. The p166-18 plasmid (~1.5-kb insertion) was sequenced using plasmid-specific and *RD3*-specific primers (Supplementary Table 3). A large part of the p417-19 plasmid (~4.5-kb insertion) was also sequenced in the same manner but the sequence failed to read through a GC-rich region of the insertion. To obtain this GC-rich sequence, the p417-19 plasmid was digested with restriction enzyme *Pvu*II and an approximately 1.1-kb fragment was subcloned into the pUC19 *Sma*I restriction site. The cloned fragment was sequenced using plasmid-specific and *RD3*-specific primers (Supplementary Table 3, Supplementary Fig. 1). The sequence of the canine *RD3* fragment corresponding to the 1552-bp-gap region in CanFam2 (CFA7: 12,832,685–12,834,236) was submitted to the NCBI database (NCBI accession No. EU687745).

#### Mutation screening in normal and affected dogs

Because of the GC-rich nature of the canine *RD3* gene (see below), genomic DNA was deaminated prior to PCR amplification using the primer pair RD3NH3F1/RD3NH3R1 (Table 2B). Deamination was undertaken using EZ Methylation-Gold Kit (Zymo Research, Orange, CA) following the manufacturer's protocol, except that the deamination step was extended to 5 h and the starting amount of DNA was reduced to 250 ng. PCR on deaminated DNA was performed under the following conditions: an initial 2-min denaturation at 96°C; then 30 cycles of 96°C (20 sec), 58°C (20 sec), and 72°C (20 sec); and a final extension step at 72°C for 5 min. PCR products were separated on 6% nondenaturing polyacrilamide gel. Eighteen *rcd2*-affected individuals, 13 *rcd2* obligate carriers, and 110 dogs not known to be affected with *rcd2* from 20 breeds were tested by PCR for the presence of the *RD3* insertion.

## Results

### Fine mapping the *rcd2* locus

#### Linkage mapping

Six individuals from the previously published *rcd2* pedigree set were recombinant in the ~4.0-Mb interval flanked by markers FH2226 and FH3972 (CFA7: 9,035,490–12,950,649 bp, CanFam2). Four of these, recombinant between FH2226 and *rcd2*, defined the centromeric end of

the zero-recombination interval. The other two, recombinant between *rcd2* and FH3972, defined the telomeric end. Genotyping the pedigrees that included these recombinant offspring with additional informative microsatellite markers CPH20, REN314O07, and FH1031 identified one individual as recombinant between CPH20 and *rcd2*, thus shifting the centromeric boundary of the zero-recombination interval and reducing this interval to approximately 2.5 Mb (CFA7: 10,014,511–12,595,819 bp, CanFam2; Fig. 1a).

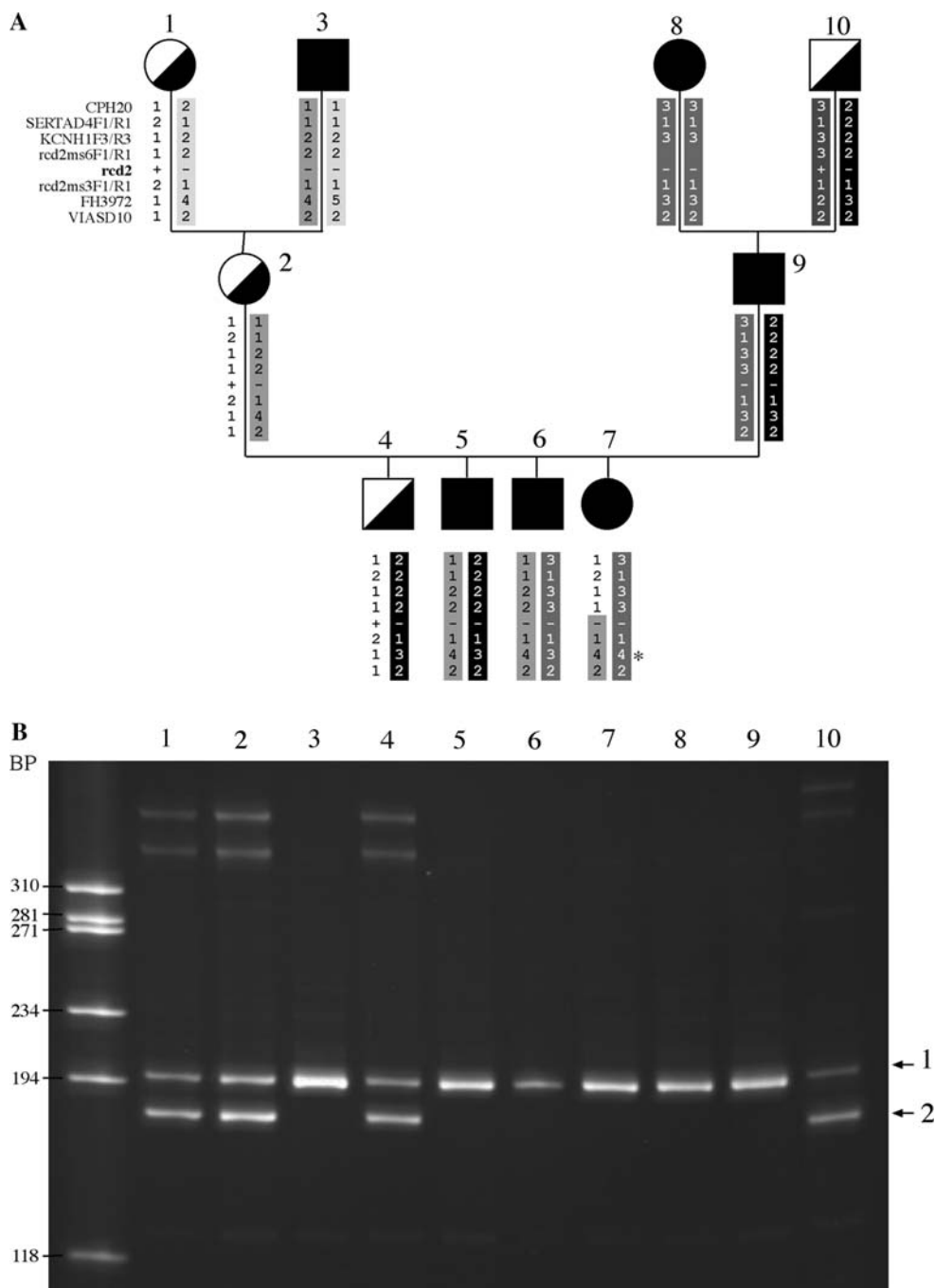
Genotyping all previously identified recombinant individuals plus 22 additional *rcd2*-informative pedigrees (140 individuals) with 6 markers (CPH20, SYT14, SERTAD4, KCNH1, FH3972, and VIASD10) identified a recombination between *rcd2* and centromeric marker KCNH1 (CFA7, Fig. 2a; Table 1), but it identified no further recombinations within the interval. Genotyping the pedigree that included this recombinant dog with another two markers (*rcd2ms3* and *rcd2ms6*) established that its recombination was between *rcd2ms6* and *rcd2* (Fig. 2a), which further reduced the *rcd2* zero-recombination interval to the approximately 308-kb distance between markers *rcd2ms6* and FH3972 (12,642,574–12,950,649 bp; Fig. 1). This reduced interval included five known genes [*RCOR3*, *GOLT1B*, *TRAF5*, *C1orf36* (*RD3*), and *SLC30A1*] on CanFam2.

#### Pedigree analysis, haplotype analysis, and LD mapping

**Pedigree analysis** Pedigree analysis was undertaken on multiple *rcd2*-affected pedigreed collie dogs that were either founders of the *rcd2*-informative collie-derived pedigrees used for linkage mapping or privately owned *rcd2*-affected pedigreed collies. From this analysis the pedigrees of purebred collies could be separated into three distinct subpopulations. In the largest of these subpopulations, the pedigrees of *rcd2*-affected collies showed a similar pattern of several inbreeding loops within recent (4–10) generations, and all these dogs shared one or another of a small number of closely related recent common ancestors that contributed to both their maternal and paternal lines of descent. As well, further descendants of these common ancestors included several additional dogs that were known historically to be affected with *rcd2*. This clearly suggested that the *rcd2* disease alleles in this subpopulation were all inherited identical-by-descent from a recent common ancestor (data not shown).

In contrast, two *rcd2*-affected pedigreed collies each demonstrated very different pedigree structures, both from each other and from the majority of *rcd2*-affected pedigreed collies (i.e., from subpopulation 1). In both of these dogs' pedigrees there was a single inbreeding loop that produced the affected dog, no further inbreeding loops were present within recent (4–6) generations, and no

**Fig. 2** Representative *rcd2*-informative experimental pedigree. **(a)** Pedigree demonstrates recombination between *rcd2* and marker *rcd2ms6*, the closest marker recombinant with *rcd2*. + indicates the wild-type allele and - indicates the *rcd2* disease allele. Haplotype shading indicates parental origin. Allele 4 for marker FH3972 in individual 7, marked by the asterisk (\*), is an example of non-Mendelian inheritance. **(b)** Cosegregation of the *RD3* insertional mutation in this *rcd2* pedigree. All affected individuals (3, 5, 6, 7, 8, 9) show one band (allele 1); all *rcd2* carriers (1, 2, 4, 10) show two bands (alleles 1 and 2)



ancestor was recognized as being historically affected by PRA. In addition, neither of these two pedigrees included any dogs identified as recent common ancestors in the first subpopulation of *rcd2* pedigrees. No ancestors were identified that were common to all three sets of pedigrees within at least 13 generations (data not shown).

Because dogs from all three subpopulations had been used in development of the *rcd2*-informative collie-derived pedigrees, it was known that the diseases segregating in each of these lines were at least allelic and probably represented inheritance identical-by-descent of a single

allele from a much more ancient ancestor common to all three subpopulations.

**Haplotype analysis and LD mapping** To identify ancestral SNP haplotypes in phase with the *rcd2* disease allele, 38 informative SNPs from the 2.57-Mb candidate region defined in Fig. 1b, with an average inter-SNP distance of 45 kb, were retrieved from the canine SNP database (<http://www.broad.mit.edu/mammals/dog/snp>) (Fig. 1b, Supplementary Table 1). The SNPs were amplified and sequenced for a subset of *rcd2*-affected and *rcd2*-

heterozygous dogs representing each of the three lines of descent described above.

Examination of the resulting SNP genotypes of *rcd2*-affected collie dogs representative of the three identified subpopulations revealed that all were homozygous throughout an extensive LD region (>2 Mb) surrounding the *rcd2* locus, but the haplotypes compared between dogs differed markedly (Fig. 1b, columns I–IV).

Dog I in Fig. 1 was homozygous for (and thus demonstrates) the haplotype characteristic of the majority of *rcd2*-affected pedigreed collies, belonging to that subpopulation described above with pedigrees characterized by inbreeding loops involving recent common ancestors. Dog II in Fig. 1 also belongs to the same subpopulation as Dog I and shares the same genotype for all markers centromeric to SNP239. At SNP239 and further telomeric SNPs, however, this dog is heterozygous, clearly presenting evidence of an historic recombination event. Dogs III and IV in Fig. 1 came from very different subpopulations, both from each other and from Dogs I and II. Both dogs are homozygous for all SNPs within the tested interval, but their haplotypes are very different, both from each other and from Dogs I and II.

The only interval in which all affected dogs shared identical alleles at all SNP loci was between SNP206 and SNP227. The boundaries of this region of common absolute LD, defined by flanking SNP196 centromerically and SNP239 telomerically, thus reduced the *rcd2* candidate interval to less than 564 kb.

#### Comparison of LD and meiotic linkage mapping results

The minimal common absolute LD region, identified above, thus exceeded 563 kb, but its centromeric boundary was more distal (telomerad) than the corresponding limit identified by meiotic linkage mapping. The 308-kb zero-recombination interval, on the other hand, established a telomeric boundary that was further proximal (i.e., centromerad) than one defined by LD mapping. The overlap of these results jointly reduced the *rcd2* candidate region to an approximately 230-kb interval bounded by SNP206 at the centromeric end (from LD analysis) and FH3972 at the telomeric end (from meiotic linkage analysis). This interval corresponded to CFA7: 12,724,381–12,950,649 on CanFam2, which included three known genes: *TRAF5*, *C1orf36* (*RD3*), and *SLC30A1*.

Analysis of *TRAF5* and *SLC30A1* genes as positional candidates genes for *rcd2*

#### *TRAF5*

An entire open reading frame of the *TRAF5* gene was amplified from the retinal cDNA of one normal and one

*rcd2*-affected dog and sequenced. Comparison of the *TRAF5* sequences detected one nucleotide difference between normal and *rcd2*-affected individuals. This polymorphism would not cause changes in the canine *TRAF5* open reading frame. Alignment of the canine and human *TRAF5* transcripts revealed 86% identity. The structure of the canine *TRAF5* ortholog did not differ significantly from that of the human or mouse gene (data not shown).

#### *SLC30A1*

Partial screening of the coding region of the canine *SLC30A1* gene (1708 bp) was performed by amplification of retinal cDNA and gDNA from normal and *rcd2*-affected individuals. Sequence comparison identified several SNPs in the coding region of the canine *SLC30A1* that were not associated with the disease (data not shown).

#### The canine *RD3* gene and transcripts

The canine gene homologous to *C1orf36/RD3* was identified as annotated but incomplete on the reverse strand of CFA7 in CanFam2 between 12,847,668 and 12,831,626. This canine assembly included a 1552-bp gap (CFA7: 12,832,685–12,834,236) downstream from the region homologous to the first coding exon of the human gene (human exon 2). Alignment of the sequence from the canine assembly with human and mouse demonstrated that the putative dog sequence appeared to be missing regions present in human and mouse and could not be fully aligned to the human coding sequence.

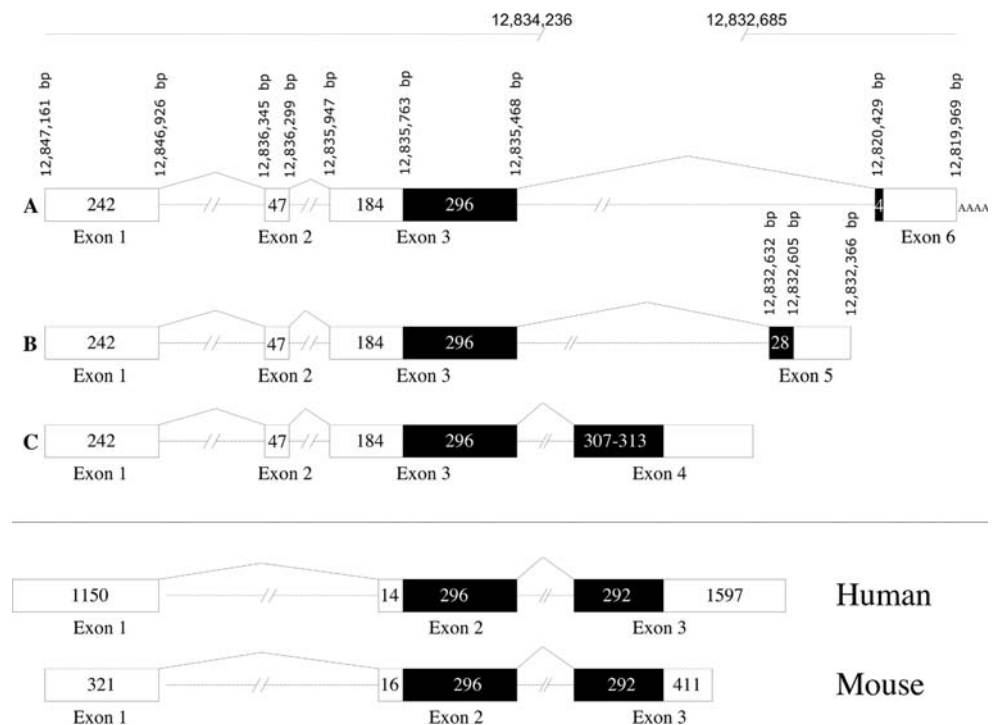
As described below, the canine gene proved, eventually, to comprise six exons that are differentially spliced to yield three splice variants (Fig. 3). Noncoding exon 1 of the dog corresponds to exon 1 of mouse and human, noncoding canine exon 2 does not appear to have a homolog in the other two species, and the first coding exon in the dog was canine exon 3 corresponding to exon 2 of the human and mouse genes. Canine exon 4 (corresponding to exon 3 of the human and mouse genes and the terminal coding exon in human and mouse) was apparently missing from the CanFam2 assembly. Canine exons 5 and 6 do not correspond to identified human or mouse exons.

#### *RD3* splice variants number 1, 2, and 3

The initial canine *RD3* transcripts were amplified and sequenced from retinal cDNA of a 10.4-week-old normal dog by 5' and 3' RACE-PCR using primers 36R1A and 36F9, respectively (Table 2A), both located in the first coding exon and designed to produce overlapping 5' and 3' RACE amplicons. The coding sequence obtained by RACE was then confirmed using RT-PCR with primers 36F18 and



**Fig. 3** Identified splice variants of canine *RD3* transcripts. Three splice variants of the canine *RD3* gene are aligned against sequence locations on canine chromosome 7 in the CanFam2 assembly of the dog genome and compared with the orthologous human and mouse transcripts. Canine exon 4 is not present in the CanFam2 assembly, as indicated by the gap from 12,834,236 to 12,832,685. Open boxes correspond to the canine *RD3* open reading frames, black boxes correspond to untranslated cDNA. (a) Splice variant #1. (b) Splice variant #2. (c) Splice variant #3. Not to scale



36R16. BLAST of the resulting canine *RD3* cDNA sequence against canine genomic DNA showed that this transcript comprised four exons which when aligned against CanFam2 were flanked by consensus splice donor and acceptor sites. This initially identified splice variant was predicted to encode a 99-amino-acid protein with a start codon located in exon 3 and a stop codon located after the first base of the fourth exon of the transcript. The predicted length of canine protein was less than that of either the mouse or the human, both of which comprise 195 amino acids. The exons comprising the dog transcript were eventually recognized as canine exons 1, 2, 3, and 6 (Fig. 3a). The transcript did not include a canine homolog of exon 3 in the mouse and human (Friedman et al. 2006) (Fig. 3, Supplementary Fig. 2) that would correspond to exon 4 in dog.

Amplification of normal retinal cDNA with primer pair 36F18 and 36JR2 identified a second splice variant (Fig. 3b, Supplementary Fig. 2). This variant included exons 1, 2, and 3 and part of a new terminating exon eventually identified as canine exon 5. Alignment against genomic sequence confirmed that the identified exons were flanked by consensus splice donor and acceptor sites. The sequence corresponding to exon 5 in this variant comprises 269 bp, but this transcript likely extends beyond the position of the primer 36JR2. This splice variant (number 2) was predicted to encode a 108-amino-acid protein with a start codon located in exon 3 and a stop codon located after the first 28 bases of exon 5.

Based on the strong conservation of the second coding exon of *RD3* among six mammalian species (Fig. 4), we

hypothesized that a canine *RD3* splice variant that includes the ortholog of this exon (exon 3 in human and mouse) should exist in the canine retinal transcriptome. We further hypothesized that absence of this transcript from the above RACE results might be caused by competition among several splice variants for shared primers and/or by the GC richness of this exon in humans and mouse (68 and 66% GC, respectively). Because this region of the canine *RD3* gene is not present in the CanFam2 assembly, a human sequence primer (36e2R2) was designed from alignment of human exon 3 and mouse exon 3 and used in conjunction with primer 36F2 (canine sequence, canine exon 1) to amplify normal canine retinal cDNA by RT-PCR. This yielded a 458-bp product that was highly similar to human and mouse *RD3* exons 2 and 3 (Fig. 4). Because this new but partial splice variant (splice variant 3) clearly established the existence of canine exon 4 (canine homolog of human and mouse exon 3), which was not present in the CanFam2 assembly, we proceeded to sequence the two plasmids p166-18 and p417-19 (Supplementary Fig. 1, Supplementary Table 3). As described in the subsection “Cloning and sequencing” in the Methods section, these plasmids were subcloned from two canine BAC clones and were expected to cover the 1552-bp-gap region in the CanFam2 assembly. Subcloning and sequencing of these genomic fragments from the BAC clones have been pursued because multiple attempts to obtain this sequence using long-range PCR have been unsuccessful. The sequence obtained from these plasmids included the newly identified canine *RD3* exon 4 and its flanking region (NCBI accession No. EU687745).

**Fig. 4** Alignment of canine *RD3* exon 4 with the corresponding exon from seven other species. Nucleotides that differ from the human sequence are colored gray. Two nucleotides in the dog sequence (positions 122–123) that serve as left and right boundaries of the insertion in the *rcd2*-affected dogs are in bold and underlined

Dog	1	GTTCCGGCAGCTGCTGGCCGAGCGCGAGCCCGAGGTGCACGAGGTGTCCGCTTCCGCTCGG
Cat	1	GTTCCGGCAGCTGCTGGCCGAGCAGGAGCCCGAGGTGCGGGAGGTGTCCGCTTCCGCTCGG
Human	1	GTTCCGGCAGCTGCTGGCCGAGCAGGAGCCCGAGGTGCAGGAGGTGTCCGCTTCCGCTCGG
Chimp	1	GTTCCGGCAGCTGCTGGCCGAGCAGGAGCCCGAGGTGCAGGAGGTGTCCGCTTCCGCTCGG
Mouse	1	GTTCCGGCAGCTGCTGGCTGAAACGAGAGCCCGAGGTGCAGGAGGTGCGCGCTGTCCGCTCGG
Chicken	1	ATTCCGGCAACTCATTGCTGAAACGAGAGCCAGAGTGCAGGAAGTATCTCGACTCTTCCGCTCTG
Opossum	1	GTTCCGGCAGCTCATTGCTGAGTACGATCCAGAAAGTCATGAGGTGTCCGCTTCCGCTTCCG
Dog	66	TGCTGCAGGAGGTCTGGAGCGCGCGCAGCGGACCGCGCCGCCCGCCCGCCCGCGGCTTGG
Cat	66	TGCTGCAGGAGGTCTGGACAGAAATGAAGCAGGAGGAGGAGGCCCGCAAGCTGACCGCCAGTGG
Human	66	TGCTGCAGGAGGTCTGGAGAGGATGAAGCAGGAAGAGGAGGCCCGCAAGCTGACCGCCAGTGG
Chimp	66	TGCTGCAGGAGGTCTGGAGAGGATGAAGCAGGAAGAGGAGGCCCGCAAGCTGACCGCCAGTGG
Mouse	66	TGTTGCAGGAGGCGCTGGAGAAATGAAGCAGGAGGAGGAGGCCCGCAAGCTGACCGCTAGTGG
Chicken	66	TTCTGCAGGAAGCTGTTGAGAAGATCAAAAGAGGAGGAAGACGCCAAGAAGCTCGCCAAAGCAATGG
Opossum	66	TCCTCCAGGAAGTCTTGAGAAGATGAAGGAAAGGAGGCTGTAGGAAGTGCACAAACAATGG
Dog	131	AGCCTGCGGCCCGCCCGCGCCCGCGCTGGACCCCGCGGGGCCCGCGCGCGCGTCTCGCCCTT
Cat	131	AGCCTGCGGCCCGCCCGCGCAACTGGCGCTGGCCACCTTCAAGACCCGCGCGCATCTCTCCCTT
Human	131	AGCCTGCGGCCCGCCCGCGCAGCCTGGCC-----ACCTTCAAGACCCGCGCGCATCTCGCCCTT
Chimp	131	AGCCTGCGGCCCGCCCGCGCAGCCTGGCC-----ACCTTCAAGACCCGCGCGCATCTCGCCCTT
Mouse	131	AGCTTGAGGCCCGCGCGCAGCCTGTCTG-----TCCTTCAAGACCCGCGCTCGCATCTCGCCCTT
Chicken	131	AACACAAAGAACAAAACAGCCTTCTCCTTAACAACGTTTAAGTCCCGGCTCAAGATCTCGCCCTT
Opossum	131	AACACTCGACCCCGGGGCACCTCTCCTGACCACTTTTAAAGACCCGTTCCCGCATCTTCTCTT
Dog	196	CGCCGGCCACATCGGCACCGTCTCCGAGGACGTGGAGCGCGACGCGCC---CGCCGCTGCGCACCT
Cat	196	CACCAGCGACATCGGACCATCTCGGAGGACGTGGAGCGGGACACCGCCGCGCGCTCCGGACCT
Human	190	CGCCAGCGACATCAGGACCATCTCCGAGGACGTGGAGCGGGACACACCGCCGCGCATGCGGTCTT
Chimp	190	CGCCAGCGACATCAGGACCATCTCCGAGGACGTGGAGCGGGACACACCGCCGCGCATGCGGTCTT
Mouse	190	TGCCAGTGACATCGAACCATCTCGGAGGATGTGGAGCGCGATGCGCCACCCCGCGCGCACCT
Chicken	196	CATCAGTGACATCAAGACCATCTCTGAGGATGTAGAGCGGGCACCAGCCACAGGAGGTTT
Opossum	196	TGCCAGTGACATCAGGACCATCTCAGAGGATGTGGAAACGGGACATCCATCCCACTCCTAGGGTCT
Dog	261	GGAGCCTGCCCCGAGTTCCGGGGCGCCGAGGGGGCGCCCGCCCGCCAGTGA
Cat	261	GGAGCCTGCCCCGAGTTCCGGGGCGCCCAAGAGGAC-----TGA
Human	255	GGAGCATGCCCGAATTCCGGGGCGCCCAAGCCGAC-----TGA
Chimp	255	GGAGCATGCCCGAATTCCGGGGCGCCCAAGCCGAC-----TGA
Mouse	255	GGAGCATGCCCGAGTTCCGGGGCGCCCGAGCCGAC-----TGA
Chicken	261	GGAGCATGCCCGAATTCCGGAATGCCAAGGATTAC-----TGA
Opossum	261	GGAGCATGCCTGAATTCCAGATGCAGAAAGCAA-----TAA

Using forward primer 36F18 (located in canine exon 1) and reverse primer 36BACR2 (located in the predicted 3' UTR downstream of the stop codon in the newly identified canine exon 4), a more complete version of splice variant number 3, which covered the entire CDS and partial 3' UTR, was amplified from normal retinal cDNA by RT-PCR. This splice variant contains exons 1, 2, 3, and 4 (Fig. 3c) and is predicted to encode either a 200- or a 202-amino-acid protein: A polymorphic repeat identified in exon 4 of several normal dogs (2 vs. 3 copies of a hexamer repeat; Fig. 5, Supplementary Fig. 2) determines the predicted deletion or insertion of two amino acids (arginine and proline).

#### Identification of the *rcd2* mutation

Attempts to amplify fragments of canine *RD3* from affected dogs were initially highly inconsistent. One primer pair (36F18, 36JR3) robustly amplified retinal cDNA from both normal and *rcd2*-affected individuals. Primers downstream from 36JR3, however, failed to amplify an affected amplicon either by RT-PCR of cDNA or by genomic PCR.

Because the normal allele of canine *RD3* exon 4 is over 80% GC-rich, we hypothesized that a mutation in the corresponding affected allele might be responsible for the difficulty in amplification of this region. To address this

possibility, we deaminated genomic DNA from normal, *rcd2*-affected, and *rcd2* carrier dogs using EZ Methylation-Gold Kit (Zymo Research). Deaminated DNA was amplified with primers RD3NH3F1/RD3NH3R1 and sequenced. Complete absence of cytosine residues in the final sequence indicated that deamination was complete. Comparison of deaminated sequences revealed a 22-bp insertion in the affected amplicon that causes a frame shift and continues the open reading frame beyond the normal stop codon.

To retrieve a further downstream sequence of the disease allele of splice variant 3, PCR was undertaken on genomic DNA from an affected dog using four primer sets (36e2F7og/36BACR2, 36e2F30/36e2R30, 36e2F30/36e2R31, 36e2F30/36e2R32). The assembled sequence from the resulting overlapping fragments included the continuation of canine exon 4 and a part of the 3' UTR of splice variant 3, but sequencing failed consistently for regions downstream of primer 36BACR2, located 39 bp downstream from the normal stop codon. Comparison of the coding sequences of *RD3* splice variant 3 from normal and *rcd2*-affected dogs predicts that the 22-bp insertion changes the last 61 amino acids of the encoded protein (Fig. 5). From analysis of the genomic DNA sequence representing the normal 3' UTR sequence for this splice variant, the affected transcript is predicted to encode a total of 574 amino acids.

```

ATGGCCCTCATCCCTGGCTTCGCTGGAATGAAGCCCCCGCGGCTGTCC
M A L I P W L R R W N E A P P R L S
CCCCGGAGGCGGCTGAGATGGTGGAGACGCTCATGATGGAGCTGGCG
P R R P A E M V L E T L M M E L A
GGGCAGATGCGGGAGGCCAGAGGCAGCAATGGGAGCGCAGTAACCGCTC
G Q M R E A E R Q Q W E R S N A L
AGGAAGTCTGCACCGCGTGGACTACAGCTGGCTGGCCAGCGCGCCGG
R K V C T G V D Y S W L A S A P R
CCCACCTACGACCTCAGCCCTGGCGAGCGGCTGCAGCTAGAGGACGTCTGT
P T Y D L S P G E R L Q L E D V C
GCCAAGATCCACCATCCTACTCGGGCCCCCCTCCTCAG\GTTCCGGCAG
A K I H P S Y C G P A I L R F R Q
CTGCTGGCCGAGCGAGCCCCGAGGTGCACGAGGTGTCGCGGCTCTTCCGC
L L A E R E P E V H E V S R L F R
TCGGTGCTGCAGGAGTCTGGAGCGCGCGCAGCGCAGCGCCGCCGCC [CGC]
S V L Q E V L E R A Q R D A A A [R]

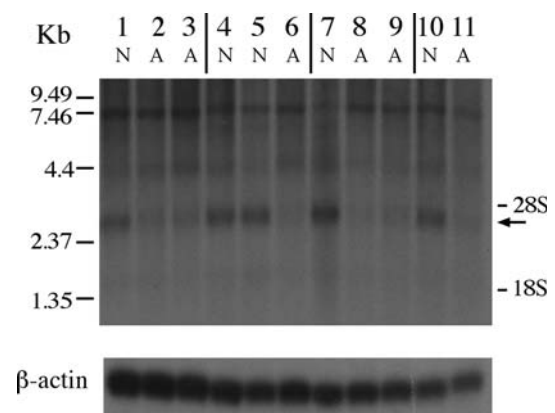
*
[CCC]2-3 C[S]G CCCTGGAGCTGCGCCCCCGCGCCCCCGCGCTGGACCCC
P ]2-3 [P/R] P W S L R P P R A P A L D P
CGCGGGCCCCGCGCGCGCTCTCGCCCTTCGCCGGCCACATCCGCACCGTC
R G P R A R V S P F A G H I R T V
GAGGACGTGGAGCGCAGCGCGCGCGCTGCGCACCTGGAGCCTGCCCGAG
E D V E R D A P P L R T W S L P E
TCCTTCGGGCGCGCAGGGGGCGCCCCGCCAGCTGA
S F R A P Q G G R P A S

```

**Fig. 5** Canine *RD3* coding sequences demonstrating variations among normal and *rcd2*-affected dogs. The nucleotide sequence of the wild-type open reading frame is shown with the corresponding translated amino acid sequence underneath in italics. The start codon and the normal stop codon are in boldface. The junction of exons 3 and 4 is represented by the backslash (G\G). A hexamer repeat is indicated within square brackets and shaded [CGCCCC]. Sequence variants with either two or three copies of this hexamer are observed from normal dogs; the *rcd2*-disease allele has two copies. This hexamer codes for a pair of amino acids (R P), which are thus predicted to be present two or three times in the translated sequence. The boxed nucleotide S represents a SNP (S = C or G), both alleles being observed in normal dogs, that changes the coded amino acid P/R from proline to arginine. The *rcd2* disease sequence has the G allele for this SNP. The asterisk (\*) marks the site (immediately before the polymorphic nucleotide indicated by S) where a 22-bp sequence (gcccgccccgccccgcccc) is inserted in the *rcd2* disease allele

### *RD3* expression in normal and mutant canine retinas

Northern analysis was undertaken to examine *RD3* expression in normal and *rcd2*-affected canine retinas at different stages of postnatal development. In normals there was an abundant transcript over 2.37 kb detected as early as 1.9 weeks of age (the earliest time point studied), the time when photoreceptors are just beginning to differentiate in the central retina; the highest level of expression was detected at 6.4 weeks of age, when retinal differentiation is near completion (Acland and Aguirre 1987; Aguirre et al. 1982) (Fig. 6). Trace amounts of a similarly sized *RD3* transcript were observed in all lanes from *rcd2*-affected retinas (Fig. 6). Less abundant signals corresponding to about 4.4 and about 7.5 kb were also observed in both normal and affected RNA samples. There was no



**Fig. 6** *RD3* Expression in canine retina. (a) Northern blot analysis of *RD3* expression in total retinal RNA from normal (lanes 1, 4, 5, 7, 10) and *rcd2*-affected (lanes 2, 3, 6, 8, 9, 11) dogs at selected postnatal age points. Lane 1 = 1.4; lanes 2 and 3 = 1.9; lane 4 = 3.3; lane 5 = 4.3; lane 6 = 4.1; lane 7 = 6.9; lane 8 = 6.4; lane 9 = 7.9; lane 10 = 9.7; and lane 11 = 10.0 weeks of age. Ribosomal RNA is indicated as 28S and 18S. (b) Northern analysis of the same RNA blot with  $\beta$ -actin probe

significant difference in intensity of these signals between normal and affected dogs. Hybridization with a  $\beta$ -actin probe was used to permit comparison of RNA loading between lanes.

### Population analysis of *RD3* mutation

To confirm that the *RD3* insertion cosegregated with the *rcd2* disease phenotype, we genotyped individuals from several *rcd2*-informative pedigrees using primers RD3NH3F1 and RD3NH3R1 (Table 2B). Complete cosegregation of the *RD3* insertion allele with the *rcd2* phenotype was observed (Fig. 2b). To test this association further, a population study was undertaken using 18 *rcd2*-affected, 13 obligate heterozygous, and 49 clinically non-affected pedigreed collie dogs that were neither part of nor closely related to the primary study population, and a further 61 clinically nonaffected dogs representing 19 other breeds (i.e., they were not rough nor smooth collies). All *rcd2*-affected dogs were homozygous for the *RD3* insertion, all obligate carriers carried both normal and affected alleles, one clinically nonaffected collie dog was heterozygous, and all other clinically nonaffected dogs tested were homozygous for the normal *RD3* allele. (Table 3, Supplementary Table 4).

### Discussion

To fine map the canine *rcd2* interval, a combination of meiotic linkage and linkage disequilibrium mapping was instrumental in reducing the candidate interval to

**Table 3** Test results of mutation screening in a population of *rcd2*-affected, carrier, and normal dogs

<i>rcd2</i> status	Affected	Obligate carriers	Clinically nonaffected collies	Clinically nonaffected dogs, other breeds
Number of chromosomes tested	36	26	98	122
Number of <i>RD3</i> insertion alleles identified	36	13	1	0

manageable proportions. This was successful in concentrating attention on the canine homolog of *C1orf36/RD3* and led to identification of the insertion proposed as the causative mutation. At the same time, this effort brought into focus critical issues in undertaking LD/association studies in a canine population.

Fine mapping by meiotic linkage analysis was undertaken using 36 mixed-breed backcross and intercross pedigrees derived from multiple *rcd2*-affected founders, including 266 offspring in the informative generation. These pedigrees were densely mapped with a set of polymorphic microsatellite markers across the previously identified *rcd2* zero-recombination interval (Kukekova et al. 2006). This effort reduced the *rcd2* interval to approximately 308 kb, with the centromeric boundary of the new interval defined by a single observed recombinant event.

Simultaneously, a SNP genotyping effort was undertaken to identify haplotypes cosegregating with the *rcd2* disease allele. The set of *rcd2*-affected individuals to genotype was selected based on ancestry analysis. The pedigrees of most *rcd2*-affected collies all shared a small number of ancestors that contributed to both their maternal and paternal lines of descent within recent (4–10) generations. Several such dogs were among the founders of the *rcd2*-informative collie-derived pedigrees used in this study. In contrast, two *rcd2*-affected pedigreed collies that were also used in the study colony were outliers. They resulted from an inbred mating but their pedigrees included none of the dogs previously identified as common ancestors in *rcd2* segregating pedigrees and, including these two dogs in the analysis, no ancestors common to all lines of descent were identified within at least 13 generations.

The SNP genotyping results were initially surprising. Comparison of genotypes representative of the majority of *rcd2*-affected pedigreed collies (Fig. 1b, Dog I) with those of the two outlier dogs (Fig. 1b, Dogs III and IV) revealed that all these dogs were homozygous for all SNPs tested over at least 2.5 Mb. However, the haplotypes differed significantly among these dogs. As fine mapping of the *rcd2* interval with SNPs continued, concentrating on the zero-recombination interval, a minimal common LD region emerged in which the same haplotype was shared by all *rcd2*-affected dogs. One *rcd2*-affected collie (Fig. 1b, Dog II), from the same subpopulation as Dog I, revealed a historic recombination event between SNP227 and

SNP239, that defined the telomeric end of the common absolute LD interval. The centromeric end of the absolute LD interval was also defined by a historic recombination between SNP196 and SNP206 (see Dog III, Fig. 1b) making the length of this interval approximately 563 kb.

Comparison of meiotic linkage analysis and LD results reduced the final *rcd2* candidate interval to about 230 kb, a region that included only three recognized genes. This demonstrates the efficiency of applying a combination of linkage and LD mapping for high-resolution mapping a monogenic trait in dogs; it also highlights the advantages and limitations of both approaches. Linkage analysis in canine populations is frequently limited by the availability of suitable informative pedigrees, particularly for purposes of fine mapping, and, in some cases, by a low level of recombination. To refine a mapped disease locus to an interval less than 1 Mb, pedigrees with well over 200 informative progeny may be required. Such pedigrees are rarely available; in the present study the pedigrees were derived from carefully bred colony dogs and could be expanded relatively rapidly as needed, but such a resource is not generally available for the majority of traits segregating in dogs. As our data show, however, once a phenotype is mapped to a broad interval, LD mapping of a few carefully selected individuals can reduce the interval dramatically, even if the disease is observed in a single breed.

The reported extensive LD within dog breeds has been broadly taken to indicate that genome-wide association studies can be undertaken at relatively low resolution within a single canine breed and then refined using the narrower LD between breeds. Indeed, this approach has had several notable successes (Goldstein et al. 2006; Karlsson et al. 2007; Parker et al. 2007; Sutter et al. 2007; Zangerl et al. 2006). However, the present study indicates that such a strategy could well fail for a disease like *rcd2* that segregates within a single breed. On the other hand, our study shows that by careful selection of dogs to genotype based on pedigree analysis, a two-stage LD analysis can be effective, even within a single canine breed. A low-resolution association study to detect LD, e.g., using currently available resources for genome-wide SNP-based association studies in dogs (Karlsson et al. 2007), would probably have failed to identify the *rcd2*-associated region if it included a mixture of dogs like Dogs I, III, and IV in Fig. 1. However, a structured association study that looked

first at the dogs with recent common ancestors and subsequently and separately at the least related *rcd2*-affected dogs would have been successful.

This observation confirms a growing recognition (Bjornerfeldt et al. 2008; Goldstein et al. 2006; Quignon et al. 2007; Parker et al. 2007) that pedigree structure within dog breeds can increase the power of association mapping studies. Several recent studies have shown that even for diseases observed in multiple breeds, the LD interval can be reduced by comparison of haplotypes of carefully selected, least related individuals from one breed to an extent comparable to that achieved by comparison of haplotypes across multiple breeds. As examples, the minimal LD interval for collie eye anomaly (CEA), a disease recognized in multiple dog breeds, was reduced to 103 kb mostly by haplotype comparison among unrelated affected Border Collies (Parker et al. 2007); the LD interval for progressive rod-cone degeneration (*prcd*), another multi-breed disease, was reducible to 106 kb largely by comparing haplotypes of unrelated Australian Cattle Dogs and Toy Poodles from different countries (Goldstein et al. 2006); and in the Golden Retriever breed, with a large population and diverse pedigree structure, selection of dogs originating from different geographical regions was able to yield a 25% decrease in LD, comparable to the LD reduction when haplotypes from two distinct breeds were combined (Quignon et al. 2007). In the current study a dramatic reduction of the *rcd2* interval was achieved with only three samples of the same geographical origin from one breed. This demonstrates that a disease segregating in just a single breed can be mapped efficiently by LD mapping, if breed size, structure, and history are appropriately taken into account. A similar mapping strategy can be predicted to be applicable for other species with well established breed structure (Amaral et al. 2008; Gautier et al. 2007; Menotti-Raymond et al. 2008; O'Brien et al. 2008).

The canine *RD3* gene has several features unlike those of the corresponding human and mouse genes. Not only are there additional exons and splice variants detected in the dog, but the extreme GC richness of canine exon 4 (corresponding to human and murine exon 3) also posed challenges for cloning and sequencing. The latter issue presumably accounts for the gap (CFA7: 12,832,685–12,834,236) in CanFam2 corresponding to this part of the genomic sequence. This sequence, eventually retrieved from BAC clones, posed even greater difficulties for screening *rcd2*-affected dogs. However, sequencing this exon from normal and *rcd2*-affected dogs after amplification of deaminated DNA identified a 22-bp insertion in the disease allele.

Three canine *RD3* retinal splice variants were identified (Fig. 3, Supplementary Fig. 2), only one of which (splice

variant 1) has a completely sequenced 3' UTR, as assessed by the identification of a poly-A tail. How far the 3' UTRs of transcripts 2 and 3 extend is unclear as attempts to RACE-PCR amplify them failed and the presently reported 3' end of the sequence for each of these splice variants is truncated by the RT-PCR primer.

Repeated attempts to sequence the disease allele of exon 4 retrieved only 39 bases downstream from the normal stop codon. Because the reading frame remained open through exon 4 to this point, it is clear that the insertion alters the predicted peptide sequence corresponding to the last 61 amino acids of the normal protein and may add a further 372 amino acids, if one assumes that the further downstream sequence of the exon 4 disease allele is the same as in the normal allele.

Northern analysis revealed one abundant *RD3* transcript, approximately 2.4 kb, in the retinal RNA of normal dogs at all developmental stages tested (Fig. 6). Only trace amounts of this transcript were detected in *rcd2*-affected individuals. Additional larger bands, about 4.4 and 7.5 kb, were also observed on Northern analysis but their intensity did not vary significantly among samples representing different developmental stages, nor between normal and affected individuals; additional studies would be required to verify which band corresponds to which splice variant. However, based on the difference between normal and affected seen in Northern analysis, the approximately 2.4-kb band probably corresponds to splice variant 3, which is the only one containing the disease-causing insertion. We hypothesize that this transcript may either be unstable in affected dogs or its high GC content makes it difficult for the cell to transcribe.

Detection of three splice variants of the canine *RD3* gene differs from reports on human and mouse *RD3* in which only a single transcript was described (Friedman et al. 2006; Lavorgna et al. 2003). The *RD3* transcript is detectable in mouse retinal RNA at a very low level at E12, an increase in expression at E18, and a further increase at P2 and P6, after which expression level remains high (Friedman et al. 2006). In normal dogs, the lowest expression level of the approximately 2.4-kb *RD3* transcript was observed at 1.4 weeks of age (the earliest time point studied), with levels increasing to 7 weeks postnatal when photoreceptor differentiation is approaching maturity in the canine retina (Acland and Aguirre 1987). The presence of several *RD3* transcripts in the canine retina may suggest that the *RD3* gene has several functions in the retina.

A C319T mutation in exon 3 of the mouse *RD3* gene, resulting in a predicted stop codon after residue 106, has been reported as the causative mutation for the *rd3* mouse phenotype (Friedman et al. 2006). Canine exon 4, mutated in *rcd2*, and mouse exon 3, mutated in *rd3*, are

orthologous. *RD3* mutations have also been reported in two human patients who are siblings and both affected with early-onset retinopathy Leber congenital amaurosis; in these cases mutations were present in the donor splice site at the end of human exon 2 (Friedman et al. 2006).

The function of the *RD3* gene is not well understood. In silico analysis of mouse and human *RD3* proteins revealed a coiled-coil domain at amino acids 22–54 and another weaker coiled-coil domain at aa 121–141. Several putative protein kinase C and consensus casein kinase II phosphorylation sites and one predicted sumoylation site have been identified in the *RD3* protein (Friedman et al. 2006). In transfected COS-1 cells, the *RD3*-GFP–fusion protein primarily exhibits nuclear or nuclear/cytoplasmic localization in close proximity to promyelocytic leukemia (PML) bodies (Friedman et al. 2006). Identification of the mutation responsible for *rcd2* reinforces the significance of the role that *RD3* plays in retinal function and development, although the specific functions need to be identified.

**Acknowledgments** We are grateful to Amanda Nickle, Gerri Antonini, and the staff of the RDS facility for research support. This study was supported by EY006855, MH077811, Foundation Fighting Blindness, Morris Animal Foundation, Collie Health Foundation, and Cornell VERGE Initiative. The content is solely the responsibility of the authors and does not necessarily represent the official views of the supporting institutions or foundations.

## References

- Acland GM, Aguirre GD (1987) Retinal degenerations in the dog: IV. Early retinal degeneration (erd) in Norwegian elkhounds. *Exp Eye Res* 44:491–521
- Acland GM, Fletcher RT, Gentleman S, Chader GJ, Aguirre GD (1989) Nonallelism of 3 genes (*rcd1*, *rcd2* and *erd*) for early onset hereditary retinal degeneration. *Exp Eye Res* 49:983–998
- Acland GM, Ray K, Mellersh CS, Gu W, Langston AA et al (1998) Linkage analysis and comparative mapping of canine progressive rod-cone degeneration (*prcd*) establishes potential locus homology with retinitis pigmentosa (RP17) in humans. *Proc Natl Acad Sci USA* 95:3048–3053
- Acland GM, Ray K, Mellersh CS, Langston AA, Rine J et al (1999) A novel retinal degeneration locus identified by linkage and comparative mapping of canine early retinal degeneration. *Genomics* 59:134–142
- Aguirre GD, Acland GM (2006) Models, mutants and man: Searching for unique phenotypes and genes in the dog model of inherited retinal degeneration. In: Ostrander EA, Giger U, Lindblad-Toh K (eds) *The Dog and its Genome*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 291–325
- Aguirre G, Farber D, Lolley R, O'Brien P, Alligood J et al (1982) Retinal degeneration in the dog. III. Abnormal cyclic nucleotide metabolism in rod-cone dysplasia. *Exp Eye Res* 35:625–642
- Amaral AJ, Megens HJ, Crooijmans RP, Heuven HC, Groenen MA (2008) Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* 179:569–579
- Bjornerfeldt S, Hailer F, Nord M, Vilà C (2008) Assortative mating and fragmentation within dog breeds. *BMC Evol Biol* 8:28
- Candille SI, Kaelin CB, Cattanaach BM, Yu B, Thompson DA et al (2007) A -defensin mutation causes black coat color in domestic dogs. *Science* 318:1418–1423
- Chader GJ, Fletcher RT, Sanyal S, Aguirre GD (1985) A review of the role of cyclic GMP in neurological mutants with photoreceptor dysplasia. *Curr Eye Res* 4:811–819
- Danciger JS, Danciger M, Nusinowitz S, Rickabaugh T, Farber DB (1999) Genetic and physical maps of the mouse *rd3* locus; exclusion of the ortholog of *USH2A*. *Mamm Genome* 10:657–661
- Friedman JS, Chang B, Kannabiran C, Chakarova C, Singh HP et al (2006) Premature truncation of a novel protein, *RD3*, exhibiting subnuclear localization is associated with retinal degeneration. *Am J Hum Genet* 79:1059–1070
- Gautier M, Faraut T, Moazami-Goudarzi K, Navratil V, Foglio M et al (2007) Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics* 177:1059–1070
- Gilbert JR, Vance JM (1994) Isolation of genomic DNA from mammalian cells. In: Haines JL et al (eds) *Current Protocols in Human Genetics*. New York: John Wiley and Sons, Appendix A.3B, pp 1–6
- Goldstein O, Zangerl B, Pearce-Kelling S, Sidjanin DJ, Kijas JW et al (2006) Linkage disequilibrium mapping in domestic dog breeds narrows the progressive rod-cone degeneration interval and identifies ancestral disease-transmitting chromosome. *Genomics* 88:541–550
- Guyon R, Lorentzen TD, Hitte C, Kim L, Cadieu E et al (2003) A 1 Mb resolution radiation hybrid map of the canine genome. *Proc Natl Acad Sci USA* 100:5296–5301
- Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC et al (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat Genet* 39:1321–1328
- Kijas JW, Cideciyan AV, Aleman TS, Pianta MJ, Pearce-Kelling SE et al (2002) Naturally occurring rhodopsin mutation in the dog causes retinal dysfunction and degeneration mimicking human dominant retinitis pigmentosa. *Proc Natl Acad Sci USA* 99:6328–6333
- Kukekova AV, Nelson J, Kuchtey RW, Lowe JK, Johnson JL et al (2006) Linkage mapping of canine rod-cone dysplasia type 2 (*rcd2*) to CFA7, the canine orthologue of human 1q32. *Invest Ophthalmol Vis Sci* 47:1210–1215
- Lavorgna G, Lestingi M, Ziviello C, Testa F, Simonelli F et al (2003) Identification and characterization of C1orf36, a transcript highly expressed in photoreceptor cells, and mutation analysis in retinitis pigmentosa. *Biochem Biophys Res Commun* 308:414–421
- Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB et al (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438:803–819
- Mellersh CS, Bourns ME, Pettitt L, Ryder EJ, Holmes NG et al (2006) Canine RPGRIP1 mutation establishes cone-rod dystrophy in miniature longhaired dachshunds as a homologue of human Leber congenital amaurosis. *Genomics* 88:293–230
- Menotti-Raymond M, David VA, Pflueger SM, Lindblad-Toh K, Wade CM et al (2008) Patterns of molecular genetic variation among cat breeds. *Genomics* 91:1–11
- Neff MW, Robertson KR, Wong AK, Safra N, Broman KW et al (2004) Breed distribution and history of canine *mdr1-1Delta*, a pharmacogenetic mutation that marks the emergence of breeds from the collie lineage. *Proc Natl Acad Sci USA* 101:11725–11730
- O'Brien SJ, Johnson W, Driscoll C, Pontius J, Pecon-Slattey J et al (2008) State of cat genomics. *Trends Genet* 24:268–279
- Parker HG, Kukekova AV, Akey DT, Goldstein O, Kirkness EF et al (2007) Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. *Genome Res* 17:1562–1571

- Quignon P, Herbin L, Cadieu E, Kirkness EF, Hédan B et al (2007) Canine population structure: assessment and impact of intra-breed stratification on SNP-based association studies. *PLoS ONE* 2:e1324
- Santos-Anderson R, Tso MO, Wolf ED (1980) An inherited retinopathy in collies: a light and electron microscopic study. *Invest Ophthalmol Vis Sci* 19:1281–1294
- Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF et al (2004) Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Res* 14:2388–2396
- Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K et al (2007) A single IGF1 allele is a major determinant of small size in dogs. *Science* 316:112–115
- Wolf ED, Vainisi SJ, Santos-Anderson R (1978) Rod-cone dysplasia in the collie. *J Am Vet Med Assoc* 173:1331–1333
- Woodford B, Liu Y, Fletcher RT, Chader GJ, Farber DB et al (1982) Cyclic nucleotide metabolism in inherited retinopathy in collies: a biochemical and histochemical study. *Exp Eye Res* 34:703–714
- Zangerl B, Goldstein O, Philp AR, Lindauer SJ, Pearce-Kelling SE et al (2006) Identical mutation in a novel retinal gene causes progressive rod-cone degeneration in dogs and retinitis pigmentosa in humans. *Genomics* 88:551–563
- Zeiss CJ, Ray K, Acland GM, Aguirre GD (2000) Mapping of X-linked progressive retinal atrophy (XLPR), the canine homolog of retinitis pigmentosa 3 (RP3). *Hum Mol Genet* 9:531–537