Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

🔓 OPEN ACCESS   Check for updates

Identification of circular RNAs hosted by the *RPGR* ORF15 genomic locus

Tatyana Appelbaum 🔘, Gustavo D. Aguirre, and William A. Beltran

Division of Experimental Retinal Therapies, Department of Clinical Sciences & Advanced Medicine, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, USA

ABSTRACT

Mutations in the retina-specific isoform of the gene encoding retinitis pigmentosa GTPase regulator (RPGR^{orf15}) cause X-linked retinitis pigmentosa, a severe and early onset inherited retinal degeneration. The underlying pathogenic mechanisms and variability in disease severity remain to be fully elucidated. The present study examines structural features of the ORF15 exonic region to provide new insights into the disease pathogenesis. Using canine and human RNA samples, we identified several novel *RPGR* ORF15-like linear RNA transcripts containing cryptic introns (exitrons) within the annotated exon ORF15. Furthermore, using outward-facing primers designed inside exitrons in the ORF15 exonic region, we found many of previously unidentified circular RNAs (circRNAs) that formed via back fusion of linear parts of the *RPGR^{orf15}* pre-mRNAs. These circRNAs (resistant to RNAse R treatment) were found in all studied cells and tissues. Notably, some circRNAs were present in cytoplasmic and polysomal RNA fractions. Although certain *RPGR* circRNAs may be cell type specific, we found some of the same circRNAs expressed in different cell types, suggesting similarities in their biogenesis and functions. Sequence analysis of *RPGR* circRNAs revealed several remarkable features, including identification of N6-methyladenosine (m6A) consensus sequence motifs and high prevalence of predictive microRNA binding sites pointing to the functional roles of these circRNAs. Our findings also illustrate the presence of non-canonical *RPGR* circRNA biogenesis pathways independent of the known back splicing mechanism. The obtained data on novel *RPGR* circRNAs further underline structural complexity of the *RPGR* ORF15 region and provide a potential molecular basis for the disease phenotypic heterogeneity.

Introduction

Retinitis pigmentosa (RP) is a heterogeneous group of inherited retinal degenerative diseases leading to vision loss [1] where X-linked RP (XLRP) comprises one of the most severe forms of the disease and accounts for 10% to 20% of all RP cases [2]. Of six disease loci mapped on the X-chromosome (https://sph.uth.edu/retnet/disease.htm) approximately 75% of XLRP cases map to the RP3 locus [3], that encodes the disease causative Retinitis pigmentosa GTPase regulator (*RPGR*) gene [4–7]. To date, more than 300 different *RPGR* sequence variants have been identified in patients with XLRP [5–7] (summarized in RetNet:https://sph.uth.edu/RetNet/). The majority of cases are associated with rod-cone dystrophy, the prevalent form of RP. Less commonly, X-linked cone-rod dystrophy, cone dystrophy and macular degeneration are caused by *RPGR* gene mutations [8–10]. Although at present there is no cure for *RPGR*-associated retinal disease, recent advances in *RPGR* research have led to the development of potential therapies, including the adeno-associated viral (AAV)-mediated *RPGR* gene replacement (https://clinicaltrials.gov/: NCT03316560, NCT03252847), following proof-of-concept studies in animal models of X-linked RP [11,12].

RPGR has a complex expression pattern and produces multiple splicing variants that widely expressed. There are two major isoforms of RPGR (RPGR^{ex1-19} and RPGR^{orf15}) found to be expressed in the retina [4,13,14]. The RPGR^{ex1-19} is encoded by exons 1 through 19 and expressed in ciliated cells throughout the body. However, to date, this transcriptional variant has not been firmly associated with any disease, although its expression levels are critical for development and maintenance of cilia length [15,16]. The RPGR^{orf15} is unique to vertebrates [17]. It's encoded by *RPGR* exons 1 through 14 and terminates with a large alternative exon ORF15 derived from the extension of exon 15 into intron 15 [5]. Interestingly, alternative splicing events (ASEs) were reported within the ORF15 region [18,19]. The RPGR^{orf15} is expressed in PR sensory cilia and basal bodies [20] and is shown to play a critical role in retinal function and viability, as multiple disease-causing mutations have been identified in *RPGR^{orf15}* in humans, dogs, and mice [5,21–23]. Of those, up to 60% occur in the *RPGR* ORF15 exon, containing a highly repetitive purine-rich internal region that encodes Glu-Gly-rich acidic domain. The repetitive nature of exon ORF15 poses challenges for molecular and clinical studies. RNA-sequencing generally fails to obtain sufficient coverage in this region [24]. Besides, cloning of the *RPGR^{orf15}* cDNA is problematic due to the instability of the ORF15 sequence, providing a basis for the use of codon-optimized constructs [25].

Extensive variability in disease phenotype is observed between patients with different ORF15 mutations, and even

---

between patients in families with the same mutation [26,27]. The molecular basis of such clinical heterogeneity in XLRP is poorly understood. It might be partially due to $RPGR^{orf15}$ allelic heterogeneity or be influenced by the action of modifier genes. In support of the latter, our recent study in canine model of $RPGR^{orf15}$-XLRP has identified several candidate genetic modifiers within the genomic region on CFA31 that could be implicated in modifying disease severity [28]. In the context of the exon ORF15 allelic heterogeneity, researches have been puzzled for years by how a fairly similar set of ORF15 mutations can result in distinct disease phenotypes. Among other factors, alternative splicing events might contribute to this effect [29,30]. Interestingly, repeat-rich sequences and non-canonical splice sites within coding exons have been shown to favour biogenesis of circular RNAs (circRNAs) [31,32], a distinct class of endogenous regulatory RNAs. CircRNAs are covalently closed RNA molecules formed by back splicing or back fusion of linear RNAs [32,33]. There is gradual recognition that many mammalian genes are able to produce circRNAs [34]. Increasing evidence suggests that circRNAs play important roles in many cellular processes, including protein translation control, gene regulation via competing with linear splicing and sponging microRNAs (miRNAs) [33–36].

Here, we aimed to examine the presence of $RPGR$ circRNAs that originate from the genomic region corresponding to exon ORF15. Our study revealed previously unknown features of $RPGR$, including the identification of a substantial number of circRNAs. The $RPGR$ circRNAs were identified in all studied cells and tissues and were present in total RNA pool as well as in cytoplasmic and polysomal RNA fractions. Furthermore, we used bioinformatic tools for sequence analysis to examine functional features of $RPGR$ circRNAs. Our data underline the complexity of the ORF15 region and provide potential molecular basis for phenotypic heterogeneity in $RPGR$ ORF15-linked disease.

Results

$RPGR^{orf15}$ and RPGR ORF15-like transcripts in canine and human cells

In canine retina RPGR^{orf15} protein is predominantly located in the connecting cilium of photoreceptors (Fig. 1A) similar to mouse and human orthologous proteins [12,20,37]. Given the lack of knowledge on whether canine $RPGR^{orf15}$ pre-mRNA produces linear ORF15-like transcripts, we analysed cDNA from canine retinal samples by RT-PCR. In normal retina we identified 3 novel $RPGR$ ORF15-like transcripts (5.4 kb (#1), 4.7 kb (#2) and 2.6 kb (#3)) (Fig. 1B, top left panel; Fig. 2A). Annotation of these transcripts showed that portions of the ORF14/ORF15 sequence were removed in a variable manner through utilization of non-canonical (atypical) splice sites. Also, retinal ORF15-like transcripts #1-3 have identical open reading frame (ORF) despite differences in the structure of 3′UTR and are predicted to encode the same protein with an estimated molecular weight 50 kDa.

The presence of $RPGR$ ORF15-like transcripts in retinal cDNAs interferes with acquiring of $RPGR^{orf15}$ transcript in RT-PCR, where $RPGR$ ORF15-like transcripts (with lesser repetitive

sequence) are preferably accumulated in the reaction. Here, canine $RPGR^{orf15}$ transcript was obtained as two overlapped RT-PCR products (encompassing exons 1–13/14 and exons 13/14-ORF15) (Fig. 1B, top left panel). Similar to mouse orthologous $Rpgr^{orf15}$ transcript [5,18], intron 14 is not spliced out from canine $RPGR^{orf15}$, forming a terminal exon composed of the exon ORF14 (exon 14 and intron 14) and the ORF15 exon. Previously reported by our group partial sequences of canine $RPGR$ (AF385629 and AF148800) support current transcript assembly (Fig. 2A, top). Western blot analysis with anti-RPGR N-terminal antibody detected three major bands (estimated as 135 kDa, 88 kDa and 50 kDa) in normal canine retinas (Fig. 1C, left panel). Of these 3 bands, we anticipate the top band (135 kDa) to be specific for RPGR^{orf15} protein as it splits in two bands with close molecular weights (135 kDa/110 kDa) in carrier females with X-linked progressive retinal atrophy 1 (xlpra1, caused by 5-bp deletion in exon ORF15 producing a premature stop resulting in a C-terminal truncation of 230 residues [22]) (Fig. 1C). Also, the lowest molecular weight band (50 kDa (Fig. 1C)) matches the molecular weight of the predicted protein encoded by $RPGR$ transcripts #1-3 (schematically shown on Fig. 2A).

We next examined the presence of $RPGR$ ORF15-like transcripts in cultured canine primary skin fibroblasts (FBs) isolated from both a wild type dog and a dog affected by xlpra2 mutation (a 2-bp deletion in exon ORF15 that induces a frameshift with inclusion of 34 basic amino acids and truncation of the terminal 161 residues [22]). We identified two $RPGR$ ORF15-like transcripts (4.3 kb (#5) and 2.6 kb (#6)) in polysome-bound mRNA fraction isolated from FBs (Fig. 1B, bottom panel). As clearly seen on Fig. 2A the structure of these two fibroblast-expressing transcripts is similar to those found in the retina. Although different in 3′UTR, these two fibroblast-expressing transcripts have the same ORF as retinal transcripts #1-3 and predict to encode the same protein. In line with this proposition, Western blot analysis identified approximately 50 kDa band in FBs (Fig. 1C, right panel). $RPGR^{orf15}$ transcript was not detected in canine primary FBs. In support, Western blot in these primary cells did not detect a band with a molecular weight corresponding to RPGR^{orf15} protein (Fig. 1C, right panel).

We further analysed cDNA samples from human retina and human cell lines ARPE19 and Y79 for the presence of $RPGR^{orf15}$ and $RPGR$ ORF15-like transcripts. Briefly, in human retina we identified $RPGR^{orf15}$ transcript (3.5 kb) and two novel $RPGR$ ORF15-like transcripts (2.6 kb (#7) and 2.1 kb (#8)) (Fig. 1B right panel, Fig. 2B). Another two novel $RPGR$ ORF15-like transcripts (2 kb (#9) and 1.6 kb (#10) were found in polysome-bound mRNA fraction isolated from ARPE19 and Y79 cells (Fig. 1B bottom panel, Fig. 2B). However, $RPGR^{orf15}$ transcript was not spotted in these human cell lines.

In conclusion, our study data support the retina-specific transcription and processing of the $RPGR^{orf15}$ transcript. We also showed that $RPGR$ genomic locus encodes various alternatively spliced $RPGR$ ORF15-like transcripts in a range of cell types. Notably, alternatively spliced introns in $RPGR$ ORF15-like transcripts are present within the region annotated as '$RPGR^{orf15}$ terminal exon'. We next sought to determine whether such internal introns (exitrons) escape degradation and become precursors of circRNAs.
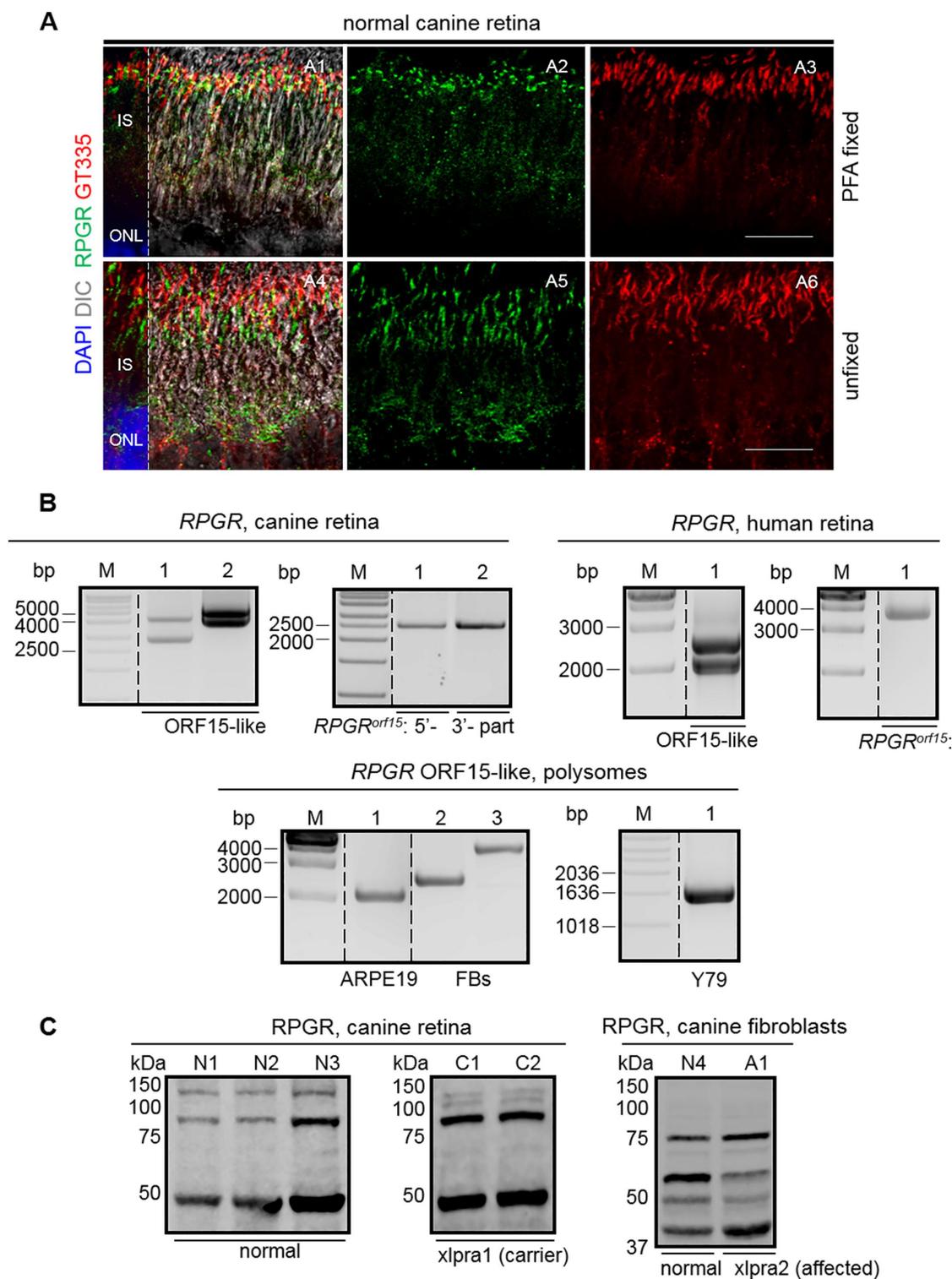
Figure 1. *RPGR* gene and protein expression in canine and human cells. Representative confocal imaging of paraformaldehyde (PFA) fixed (A1-A3) and unfixed (A4-A6) normal canine retinas immunostained with anti-RPGR (green) and ciliary marker polyglutamylation modification GT335 (red) antibodies. We found that PFA fixation of the retinal tissue interferes with RPGR labelling of photoreceptors (the signal was obtained only from few retinal OCT sections); therefore, the results were confirmed in unfixed OCT retinal sections. In photoreceptor cells, connecting cilia and to a certain extent the downstream area towards the inner segment display strong RPGR labelling in both PFA fixed and unfixed retinal sections. *Note*: ONL = outer nuclear layer; IS = inner segment of photoreceptors. A1 and A4 show the merged image for the three channels, including DAPI nuclear stain on the left of each set, separated by dashed line. A2, A3, A5 and A6 images are presented as an individual colour to better illustrate a particular feature. Scale bar = 10 μm. (B) RT-PCR expression analysis of *RPGR^{orf15}* and *RPGR* ORF15-like transcripts in canine (*top left*) and human (*top right*) retinas. (*bottom*) *RPGR* ORF15-like transcript were also identified in polysome-bound mRNA fraction isolated from human cell lines (ARPE19 and Y79) and cultured canine primary skin fibroblasts (FBs) isolated from both xlpra2 mutant and normal dogs (*bottom left*, lane 2 and 3, respectively). *Note*: The irrelevant lanes were spliced out from the 12-wells agarose gel (indicated by dashed lines). (C) Western blot analysis of RPGR protein in canine retinas and primary skin fibroblasts.
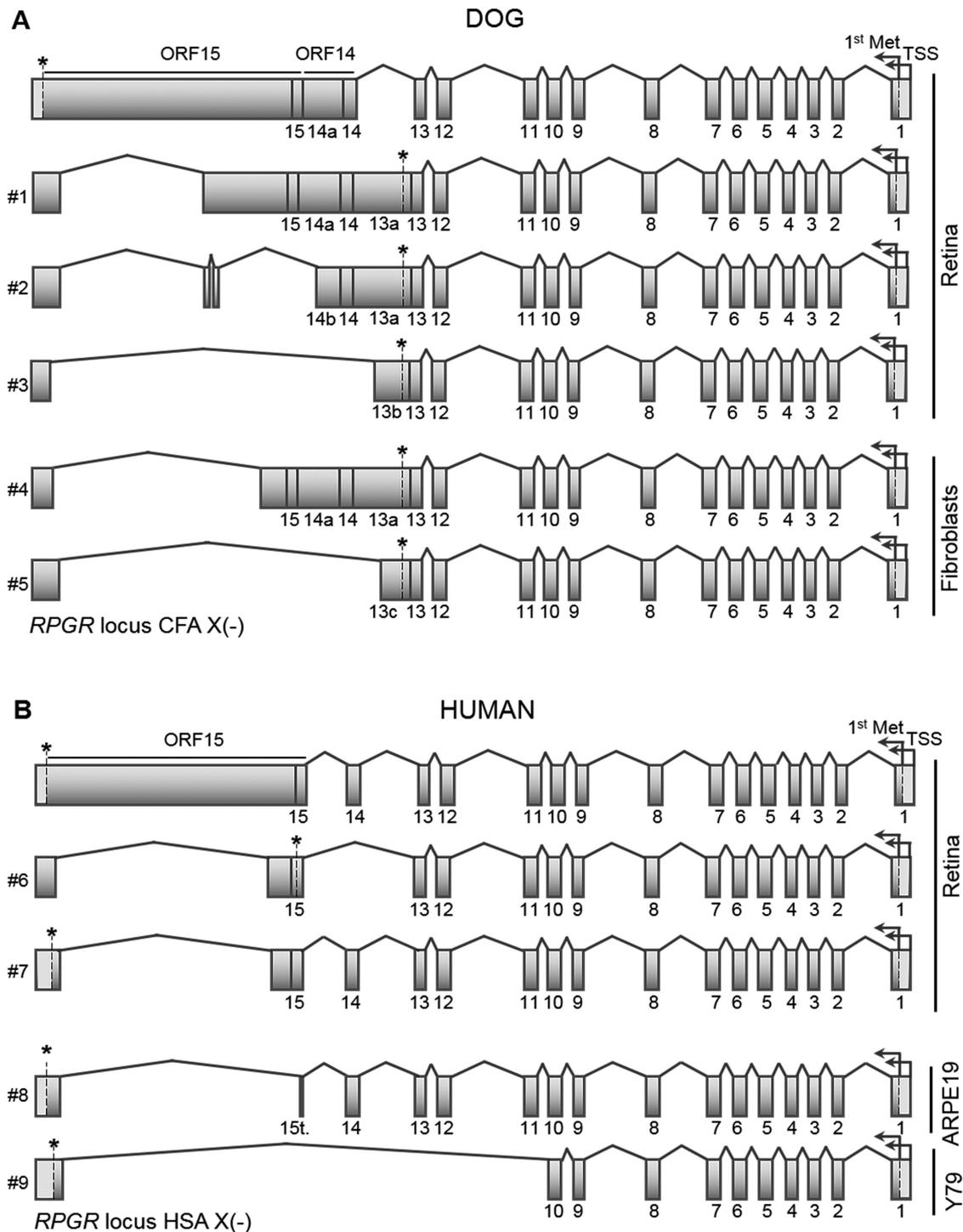
Figure 2. Structural organization of *RPGR^orf15^* and *RPGR* ORF15-like transcripts. Exon structure of the *RPGR^orf15^* and alternative transcripts in (A) canine and (B) human cells. *Note*: Exons are numerated, and predicted stop-codon is indicated by asterisk. TSS = transcription start site. GenBank Accession Numbers for the sequences are listed in Supplementary Table S1.

Numerous circRNAs are hosted by RPGR exon ORF15 region

Here, we tested the hypothesis that the repetitive region of the *RPGR^orf15^* terminal exon (that has a high rate of ASEs) is prone to form circRNAs. Towards this end, we performed RT-PCR in cDNA transcribed from total RNA samples isolated from canine (retina, brain and FBs) and human (retina, ARPE19 and Y79)

cells using outward-facing primers designed inside identified exitrons in ORF14/ORF15 (dog) and ORF15 (human) exonic regions (as shown on Fig. 3A and 3E, respectively). We obtained a subset of RT-PCR products where 3'- and 5'-part of linear *RPGR^orf15^* (derived from forward and reverse primers, respectively) were joined accurately, supporting back fusion events for
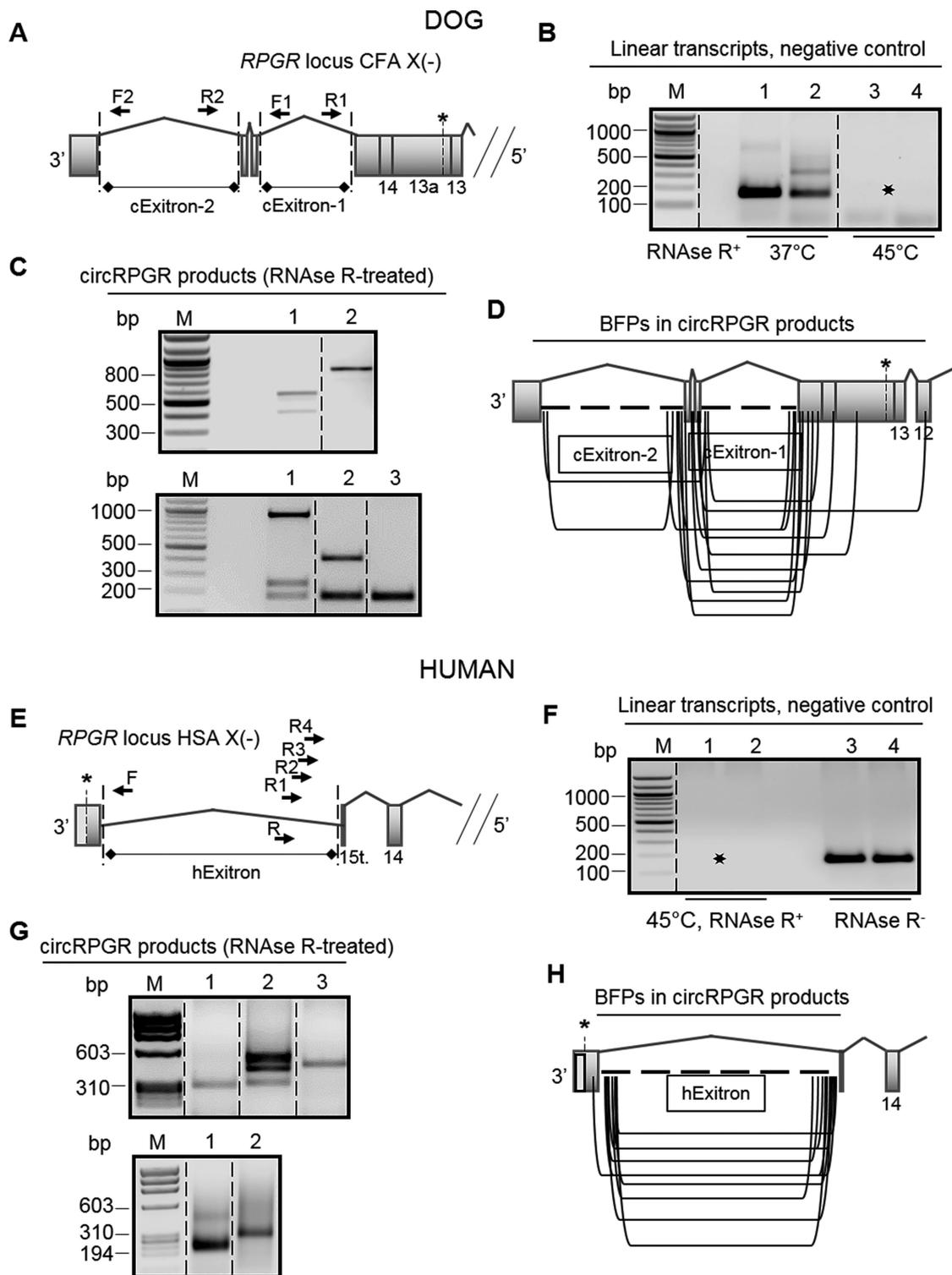
Figure 3. Identification of *RPGR* circRNAs in canine and human cells. For circRNA detection outward-facing primers were designed inside identified exitrons in canine ORF14/ORF15 (A) and human ORF15 (E) exonic regions. Reaction conditions for RNAse R treatment were optimized using canine retinal RNA samples and reactions were carried at 37°C (2.5 hours) and 45°C (2 hours) (B). The optimized conditions for RNase R treatment (45°C, 2 hours) were also applied to human RNA samples (F). Representative RT-PCR images of circRNA products obtained in RNase R-treated canine (C) and human (G) RNA samples. *Note*: The irrelevant lanes were spliced out from the 12-wells agarose gel (indicated by dashed lines). The location of identified back fusion points (in linear format) is schematically shown for canine (D) and human (H) sequences.

the formation of these products. Subsequently, the circularity of these candidate circRNAs was confirmed as described below.

It has become common to validate *bona fide* circRNAs by confirming that they are resistant to degradation by the 3'-5'

exoribonuclease Ribonuclease R (RNase R) [38]. However, although RNase R efficiently degrades linear RNAs, some linear transcripts were shown to be resistant to RNase R [39]. Therefore, using canine RNA samples we first tested

whether linear *RPGR* transcript can be efficiently degraded by RNase R. Surprisingly, *RPGR* failed to be fully digested at standard enzymatic conditions at 37°C (Fig. 3B). Replacement of K+ with Li+ in the RNase R reaction buffer that was previously reported to be helpful for digestion of highly structured RNA [39], was not effective (data not shown). We overcame this challenge by increasing the reaction temperature to 45°C (Fig. 3B). Using such optimized conditions for RNase R treatment we similarly treated canine (retina, brain and FBs) and human (retina, ARPE19 and Y79) RNA samples to enrich circular *RPGR* (circRPGR) RNAs (exemplified on Fig. 3C and 3G). Further RT-PCR and sequencing of PCR products obtained from RNase R-treated samples identified 20 canine and 16 human circRPGR sequences derived from the $RPGR^{orf15}$ terminal exon region (see Table 1, Fig. 3D, 3H, Figs. 4 –5 and Supplementary Fig. S1). For simplicity, we mapped the identified back fusion points (BFPs) in circRPGRs relatively to $RPGR^{orf15}$ reference coding sequence (canine OP555994 and human NM_001034853, respectively). Briefly, 7 circRPGR products (3 canine and 4 human) were detected in the cytoplasm of the cultured cells and 8 circRPGR RNAs (7 canine and one human) were present in polysome fraction. Remaining

sequences were obtained from cDNA transcribed from total RNA.

We next organized circRPGR sequences in four groups (I–IV) based on their distinct structural features. Two canine circRPGRs (cfa-circRPGR-11 and cfa-circRPGR-12) identified in cytoplasmic fraction of canine FBs, were assigned to group I 'multi-exonic' (Fig. 4A). CircRPGR products derived solely from the region corresponding to the $RPGR^{orf15}$ terminal exon were assigned to group II ('sub-exonic', n = 28) (exemplified on Fig. 4B and 5A). Notably, 3 members of group II (cfa-circRPGR-01, cfa-circRPGR-10 and cfa-circRPGR-19) were found in more than one cell type (see Table 1), which suggests their potential functionality as well as similarities to the pre-mRNA processing. Next, 3 circRPGR products (cfa-circRPGR-03, cfa-circRPGR-15 and hsa-circRPGR-03) that were structurally similar to group II circRPGRs but also contained partial intronic upstream sequence (exemplified on Fig. 5B) were added to group III (exonic-intronic). Lastly, 3 circRPGR products (cfa-circRPGR-17, hsa-circRPGR-10 and hsa-circRPGR-15) were assigned to group III 'shuffled sequence'. These circular transcripts are non-linear chimeric molecule that are topologically inconsistent with the reference sequence (as reported in some other circRNAs [40]). Potentially these

Table 1. Annotation of circRPGR products.

study ID	gr.	BFP 3'_5' (cDNA pos.) *	D-N	SHSs	cells/tissue (***)
				motifs adjacent to the BFP **	
cfa-circRPGR-01	II	c.2586_c.1973	ag/ca	-	retina (p), FBs (t), brain (t)
cfa-circRPGR-02	II	c.2615_c.1992	-	gaggcaaagtga/gagacaaagtga	retina (t), retina (p)
cfa-circRPGR-03	III	c.2669_int13.2024	gg/tc	-	retina (t)
cfa-circRPGR-04	II	c.2667_c.1992	ag/ac	-	retina (t)
cfa-circRPGR-05	II	c.2577_c.1966	ag/ga	-	retina (t)
cfa-circRPGR-06	II	c.3772_c.2528	ag/ac	-	retina (t)
cfa-circRPGR-07	II	c.3744_c.2767	-	gaaggagagga/gaaggagagga	retina (t)
cfa-circRPGR-08	II	c.2740_c.1902	tc/ag	-	retina (p)
cfa-circRPGR-09	II	c.2713_c.1991	gg/ga	-	retina (p)
cfa-circRPGR-10	II	c.2791_c.2066	-	gatga/gatga	retina (t, p), FBs (t)
cfa-circRPGR-11	I	c.2682_c.1445	-	aagggagagaaa/aagggaaagaaa	FBs (c)
cfa-circRPGR-12	I	c.3472_c.1960	ag/ct	-	FBs (c)
cfa-circRPGR-13	II	c.2749._c.1904	aa/gc	-	FBs (c)
cfa-circRPGR-14	II	c.3686_c.1659	gg/ga	-	FBs (t)
cfa-circRPGR-15	III	c.3445_int13.1689	ag/aa	-	FBs (t)
cfa-circRPGR-16	II	c.2794_c.2003	-	gagaa/gagaa	FBs (t)
cfa-circRPGR-17	IV	c.2475_c.1977	-	gcctgcagg/gtctgcagg	FBs (t)
cfa-circRPGR-18	II	c.2661_c.1969	-	gaaatgg/gaaatgg	FBs (p)
cfa-circRPGR-19	II	c.2760_c.1911	ga/tc	-	retina (p), FBs (p)
cfa-circRPGR-20	II	c.2734_c.1904	-	aggcca/aggcca	brain (t)
hsa-circRPGR-01	II	c.3054_c.1992	aa/ag	-	ARPE19 (t)
hsa-circRPGR-02	II	c.3178_c.1778	-	gagaagagga/gagcagagga	ARPE19 (c)
hsa-circRPGR-03	III	c.3316_int14.586	-	aaagt/aaagt	ARPE19 (c)
hsa-circRPGR-04	II	c.3164_c.2004	aa/ct	-	Y79 (t, p)
hsa-circRPGR-05	II	c.2489_c.2068	-	aggag/aggag	Y79 (t)
hsa-circRPGR-06	II	c.3163_c.1826	aa/at	-	Y79 (t)
hsa-circRPGR-07	II	c.3092_c.1860	ga/gg	-	Y79 (t)
hsa-circRPGR-08	II	c.3078_c.1814	ag/ag	-	Y79 (t)
hsa-circRPGR-09	II	c.3018_c.1853	-	gaggaagaa/gaggaagaa	Y79 (t)
hsa-circRPGR-10	IV	c.2385_c.1933	ag/gg	-	Y79 (c)
hsa-circRPGR-11	II	c.3236_c.1777	ag/gc	-	Y79 (c)
hsa-circRPGR-12	II	c.2990_c.2217	-	agaagaag/agaagaag	retina (t)
hsa-circRPGR-13	II	c.3143_c.2196	-	aggag/aggag	retina (t)
hsa-circRPGR-14	II	c.3021_c.2136	-	agaagg/aaaagg	retina (t)
hsa-circRPGR-15	IV	c.2320_c.1816	aa/gt	-	retina (t)
hsa-circRPGR-16	II	c.3183_c.1768	-	gaggagga/gagaagga	retina (t)

Note: * Unless intronic, BFPs (composed from 3' and 5' fusion nucleotides) are mapped relatively to $RPGR^{orf15}$ reference cDNA sequence (OP555994 (dog) and NM_001034853 (human), respectively). ** Dinucleotides (D-N) and short homologous sequences (SHSs) are shown relatively to their position on cDNA (3'/5'). Back fusion nucleotides in dinucleotide flanking motifs and candidate back fusion nucleotides within SHS flanks are marked in bold. Mismatched nucleotides within the SHSs are shown in italics. *** cDNA source: (t) total RNA, (p) – polysome fraction and (c) cytoplasm.
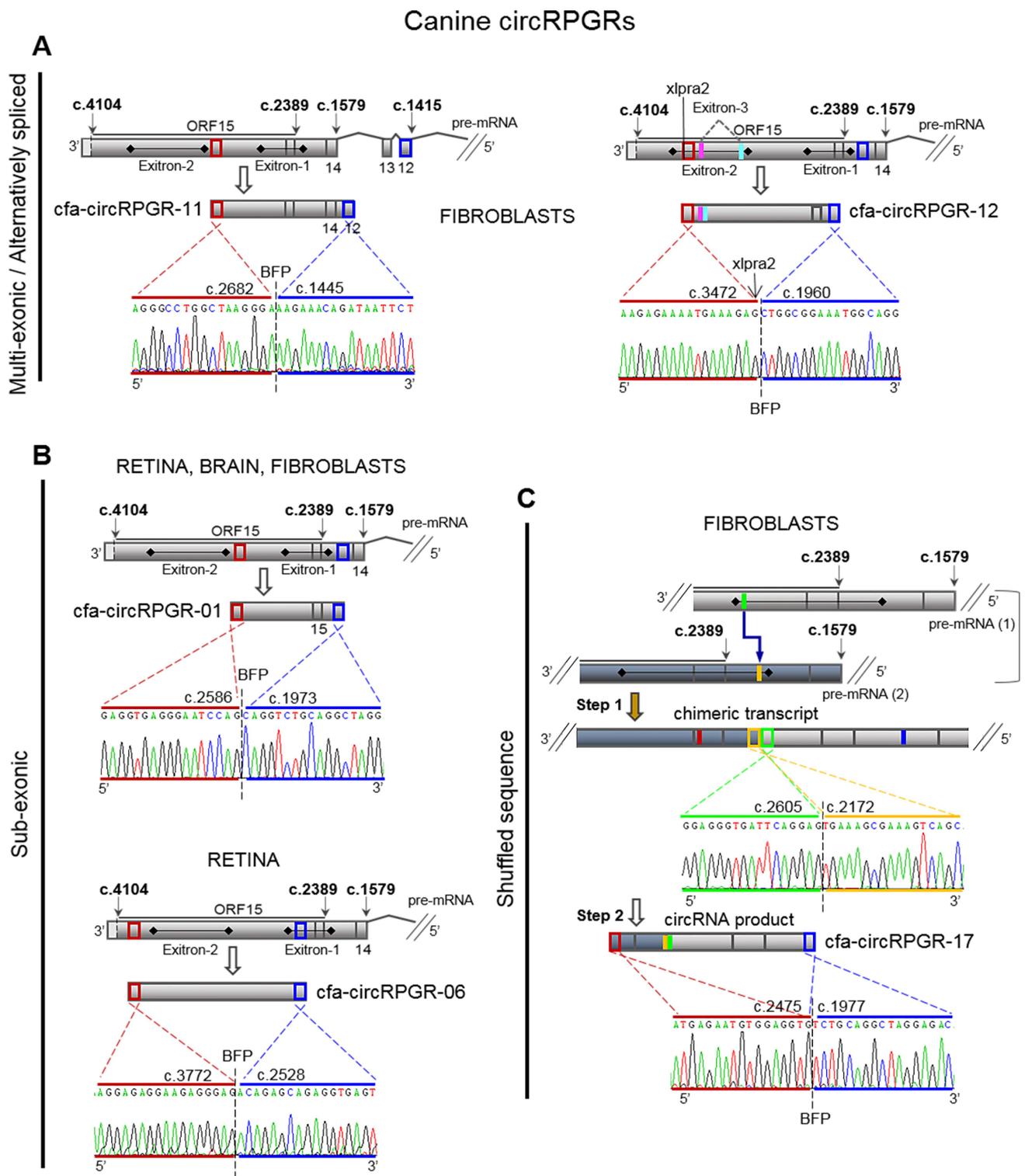
Figure 4. Structural features canine *RPGR* circRNAs. (A) group I (Multi-exonic) is represented by two fibroblast-expressed circRNAs that have alternatively spliced sequences. (B) the structure of the group II (Sub-exonic) representatives, cfa-cirRPGR-01 and cfa-cirRPGR-06. (C) the structure of the group IV (Shuffled sequence) representative cfa-cirRPGR-17 and suggested steps in its biogenesis, including (step 1) formation of a chimeric transcript formed by two different linear *RPGR* transcripts (colour-coded) and (step 2) back fusion events. *Note*: For simplicity the structure of circRPGR products is shown in linear format.

circular transcripts are excised from a chimeric *RPGR* transcript formed by non-canonical splicing events (such as *trans*-splicing) between two linear *RPGR* transcripts (schematically shown on Fig. 4C and Fig. 5C).

Detailed sequence analysis of circRPGRs revealed the complexity of alternative splicing events in this *RPGR* region. We identified short homologous sequences (SHSs) adjacent to the 5'-end and 3'-end BFP (relatively to parental linear RNA

Figure 5. Structural features human *RPGR* circRNAs. (A) group II (Sub-exonic) is represented by the structure of 3 circRNAs found in human cultured cells. (B) group III (Exonic-intronic) representative hsa-cirRPGR-03. (C) the structure of the group IV (Shuffled sequence) circRNA hsa-cirRPGR-10 and hsa-cirRPGR-15 as well as suggested steps in their biogenesis, including (step 1) formation of a chimeric transcript formed by two different linear *RPGR* transcripts (colour-coded) and (step 2) back fusion events.

sequence) in approximately half of circRPGRs (40% canine and 50% human) sequences (Table 1). SHS pairs were represented by identical direct repeats (11 of 16 total cases) or nearly identical direct repeats (with a single mismatch in 3'-end SHS (5 of 16 total cases)) of lengths between 5- to 12-nt. Further examination of nearly identical SHS pairs showed that

mature cirRPGR sequences contain the SHS sequence without the mismatch. Previous studies showed that SHS pairs can contribute to production of natural chimeric transcripts through template switching in RNA processing [41,42] as well as may assist the biogenesis of interior circRNAs [32]. The potential SHS-based model for circRPGR biogenesis is

shown in Supplementary Fig. S2 using hsa-circRPGR-09 and cfa-circRPGR-02 sequences as an example. Interestingly, group II member cfa-circRPGR-02 is a representative of so-called 'intron circles', that contain the entire intron sequence. To our knowledge, only few intron circles in other genes have been reported to date [43,44]. Here, the sequence analysis revealed that cfa-circRPGR-02 contains the entire exitron excised from one of newly identified canine *RPGR* ORF15 transcripts (MK453375).

The remaining circRPGRs which lack SHSs at proximity to BFP, appear to contain diverse dinucleotide motifs adjacent to the BFP (Table 1), suggesting an alternative mechanism (so far unidentified) in biogenesis of such circRPGRs. Although AG motif adjacent to the 3' back fusion nucleotide was the most representative (found in 10 of 20 total sequences), other motifs varied significantly. Thus, our data point to the presence of two different non-canonical pathways in circRPGR biogenesis (SHS-based and SHS-independent routes), although the precise mechanisms have yet to be defined.

Impact of natural sequence variants on biogenesis of circRPGRs

Sequence variants in splice sites and/or in proximal *cis*-regulatory sequences that occur naturally or due to mutagenesis approach can result in splicing alteration in linear and circular transcripts [41,45–47]. Taking these data into consideration we sought genetic variants within ORF15 exonic region that may impact biogenesis of circRPGRs. Although we didn't perform mutagenesis experiments, we used sequence analysis to evaluate the likelihood of BFPs disruption. Briefly, we found that the BFP (c.3472_c.1960) in cfa-circRPGR-12 (expressed in normal canine primary fibroblasts) can be influenced by the xlpra2 mutation (c.3472_3473del) (Fig. 4A). To examine this possibility, we carried out RT-PCR in cDNA from xlpra2 mutant canine fibroblasts using RT-PCR primers designed in a proximity to the BFP of interest. We did not obtain PCR product containing this BFP likely due to the lack of the expression of this circRNA. We also identified sequence variant rs2067186077 (c.2320 G > A) in the hsa-circRPGR-15 BFP (c.2320_c.1816) (Fig. 5C). It cannot be excluded that this is a cryptic BFP that was formed due to the presence of rs2067186077 in this position. We further explored whether any reported sequence variants in human ORF15 region are positionally overlapped with circRPGR BFPs. For this, we queried the dbSNP and the LOVD database and found that identified BFPs can be potentially influenced by natural sequence variants including deleterious mutations (see Supplementary Table S2 for details). These data suggest that certain sequence changes within exon ORF15 might lead to the defects in circRPGRs biogenesis.

Circular RPGR products possibly affecting cellular pathways

The presence of circRPGR products in the cytoplasm and polysomes suggests their functional significance. Here, using a bioinformatic approach we examined the potential for circRPGRs translation and their ability to target miRNAs.

It was previously shown that some circRNAs can be translated into proteins in a cap-independent and internal ribosome entry site (IRES)- or N6-methyladenosine (m6A)-dependent manner [35,48]. According to a recent study, m6A, the most abundant base modification of RNA, promotes efficient initiation of protein translation from circRNAs [35]. Sequence analysis of 7 canine circRPGR products and one human circRPGR detected in polysome fraction showed the presence of m6A consensus motifs ('RRm6ACH', R = G or A; H = A, C or U [35]) in 5' proximity to the AUG codons. Specifically, 7 canine polysome-bound circRPGR RNAs (cfa-circRPGR-01, cfa-circRPGR-02, cfa-circRPGR-08, cfa-circRPGR-09, cfa-circRPGR-10, cfa-circRPGR-18 and cfa-circRPGR-19) are formed within ORF14/ORF15 exonic region corresponding c.1902-c.2793 of the reference sequence (OP555994). This region contains 5 AUG codons in positions c.1969 (p. Met657), c.2023 (p.Met675), c.2302 (p.Met768), c.2350 (p. Met784) and c.2662 (p.Met888). m6A consensus motifs were identified in positions c.1934–1938, c.2009–2014, c.2277–2281, c.2329–2333 and c.2643–2647. The sequence of human polysome-bound circRPGR RNA hsa-circRPGR -02 is derived from ORF15 exonc region corresponding c.2004-c.3164 (NM_001034853), containing two AUG codons in positions c.2057 (p.Met686) and c.2179 (p. Met727). Similar to dog, m6A consensus motifs in human sequence were found upstream to methionine positions in c.2037-c.2041 and c.2169–2173, respectively. Predicted canine and human circRPGR ORFs appear to be in frame with the reference proteins but differ in their C-terminal part (schematic representation of these ORFs shown on Fig. 6). Briefly, we found that 5 out of a total of 8 circRPGRs (cfa-circRPGR -02, cfa-circRPGR-09, cfa-circRPGR-10, cfa-circRPGR-18 and hsa-circRPGR-04) has a potential to be translated by a rolling circle amplification mechanism [49], initiated from either methionine. In the remaining 3 circRPGR products (cfa-circRPGR-01, cfa-circRPGR-08 and cfa-circRPGR-19) circular junctions create frameshifting until it reaches a stop codon present in the new frame.

It is known that a subset of naturally expressed circRNAs can bind to miRNAs to suppress their function, acting as competitive endogenous RNAs [50]. In fact, the ability to functionally sequester endogenous miRNAs is the most commonly reported function of circRNAs. Here, we sought to examine whether the cytoplasmic circRPGR products have a potential to target miRNAs. Each of circRPGR transcripts contains a significant portion of *RPGR*^{orf15} terminal exon. By querying the miRDB database for miRNA target prediction we found that the sequence within canine and human *RPGR*^{orf15} terminal exon contains multiple potential miRNAs binding sites. Top results with the highest target score are shown in Supplementary Tables S3 and S4. Briefly, 92 binding sites for 10 miRNAs were identified in canine ORF14/ORF15 region. As the human genome is better annotated comparatively to that of the dog, we retrieved 636 binding sites for 35 miRNAs in human *RPGR* ORF15 sequence. None of these miRNAs binding sites has much significant complementarity beyond the seed region, indicating that these miRNAs may bind but likely cannot slice circRPGRs. Hence, these circular transcripts might function as efficient miRNA sponges.
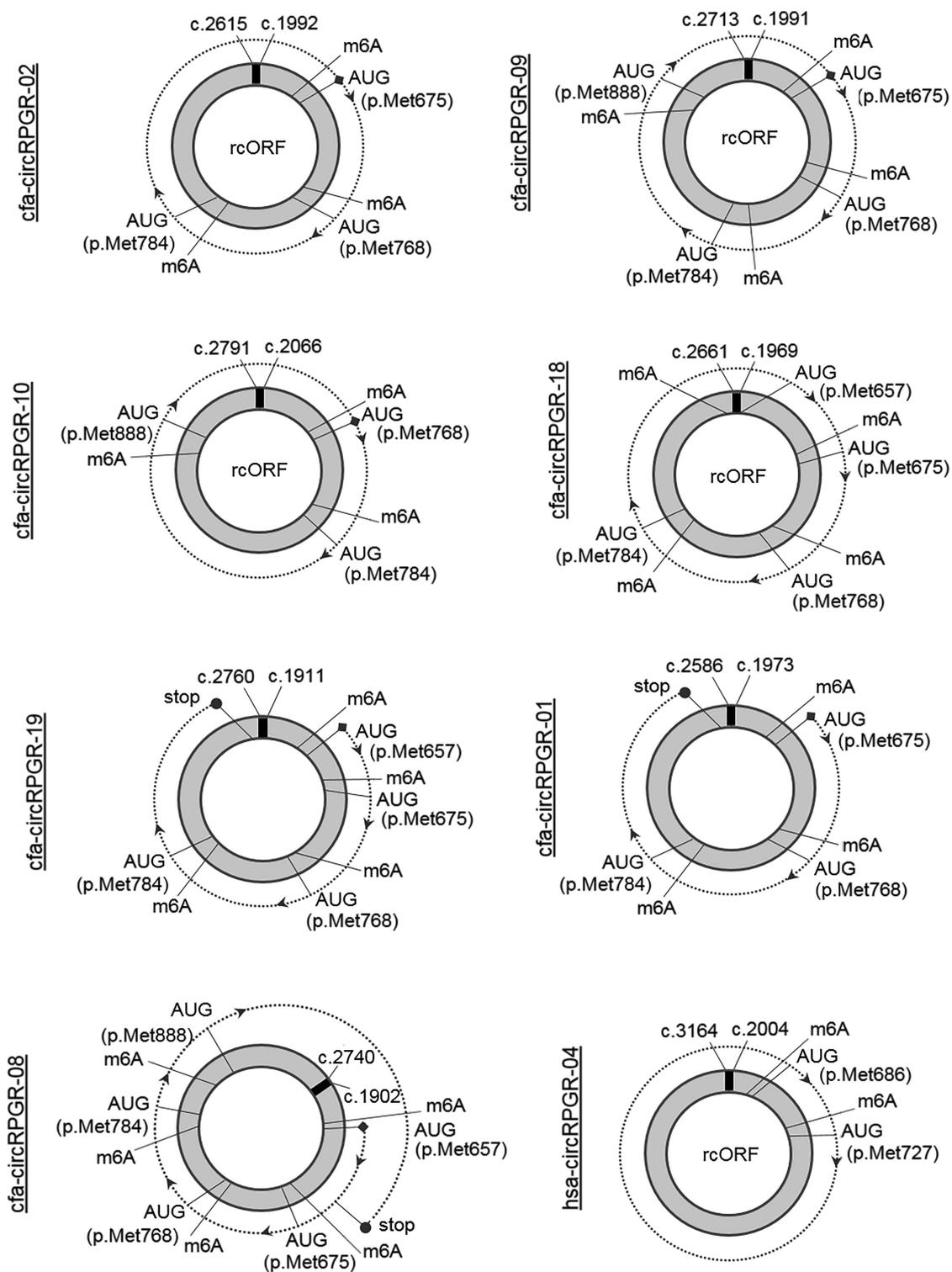
Figure 6. Predicted ORF structures in canine and human *RPGR* circRNA sequences. Sequence analysis of 7 canine and one human circRNAs identified in polysome-bound mRNA fractions indicated potential ORF (schematically shown) within the sequences. *Note*: rcORF = rolling cycle ORF. Where applicable, predicted stop-codon is marked by bulb.

Overall, annotation of canine and human circRPGR sequences assigned putative functions to a subset of *RPGR* circular products, demonstrating the capacity of *RPGR* ORF15 genomic region to produce different gene products that may serve multiple tasks.

LncRNA ROBO2-AS as a candidate modulator of RPGR^{orf15} pre-mRNA processing

Alternative splicing of *RPGR*^{orf15} pre-mRNA produce linear *RPGR*^{orf15} and *RPGR* ORF15-like transcripts as well as circRPGRs suggesting that the pre-mRNA processing is

under tight regulation. Unlike *RPGR* ORF15-like and circRPGR transcripts that were present in all cell types tested, expression of *RPGR*^{orf15} is limited to the retina, likely being photoreceptors specific. On the basis of these observations, we speculate that certain retinal non-coding RNAs may be involved in splicing control of *RPGR*^{orf15} terminal exon perhaps by forming RNA-RNA duplexes. Such RNA-RNA duplexes between sense and antisense transcripts known to modulate binding of splicing regulatory proteins as most of them bind to single-stranded RNA [51]. To further test this possibility, we examined the presence of natural RNA-RNA duplex between *RPGR*^{orf15} and lncRNA *ROBO2-AS* in canine retina. This lncRNA was identified in our previous screen as a candidate genetic modifier for *RPGR* retinal degeneration in canine model of *RPGR*^{orf15}-XLRP [28]. *ROBO2-AS* transcriptional variants contain exons that lack homology to its natural *cis* target *ROBO2* mRNA [28]. We assumed that *ROBO2-AS* may also act as *trans*-antisense RNA and be involved in transcriptional and/or translational control of other genes.

Computational modelling with the IntaRNA software, which considers both Watson-Crick (G-C, A-U) and wobble (G-U) base pairing, predicted interaction sites in both RNA sequences. Specifically, a significant part of the *ROBO2-AS* exon 3 (pos. 602–843 in MK450419) is predicted to have multiple overlapping interaction sites in canine exon ORF15 region (c.2389-c.3950) with the minimal energy of RNA-RNA interaction (E^{min}) varied between −56.8 kcal/mol to −72.8 kcal/mol (Fig. 7A). No RNA-RNA interactions were predicted between *ROBO2-AS* (exon 3) and *RPGR* exons 1–14 (E^{min} didn't exceed −20.4 kcal/mol), pointing to the likelihood that the *RPGR* exon ORF15 could be a selective *trans*-antisense target for *ROBO2-AS*. Interestingly, the *ROBO2-AS* exon 3 contains the sequence variant that showed strong association with milder disease phenotype [28]. Also, although expression of the *ROBO2-AS* exon 3 was detected in canine retinal tissue, it was not found in primary fibroblasts as is the case with *RPGR*^{orf15}.

Trans-antisense RNAs generally display more limited complementarity to their target mRNA that do *cis*-antisense RNAs [52]. Consistent with this, predicted *ORF15/ROBO2-AS* RNA-RNA duplexes contain non-contiguous segments that would complicate the performance of a ribonuclease protection assay. Therefore, we leveraged a biotin-streptavidin affinity purification method to examine this interaction. Briefly, a set of short 5'-biotin-labelled oligonucleotides (15–17 nt) targeting *RPGR* mRNA (exons 3, 7 and 12) were used to isolate *RPGR* bound to streptavidin magnetic beads in total RNA samples extracted from normal and xlpra1 mutant dog retinas (24 wks and 16 wks old, respectively). Subsequent RT-
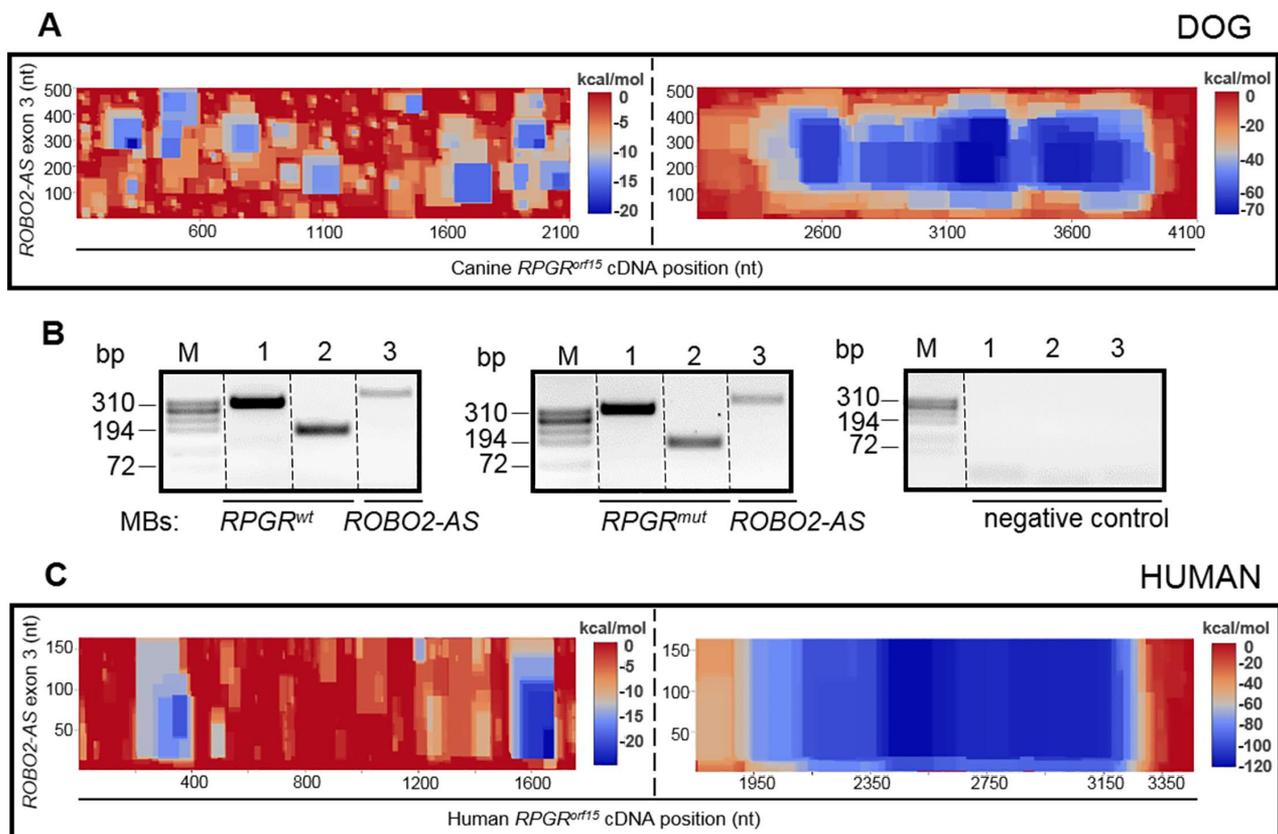

Figure 7. *RPGR*^{orf15} mRNA/*ROBO2-AS* (exon 3) interaction abundance. (A (dog) and C (human)). The plot (in a heatmap style) visualizes the minimal energy (kcal/mol) of any RNA-RNA interaction that can be formed between these two RNAs. (B) RT-PCR results in canine cDNA (generated from *RPGR* RNA bound to magnetic beads (MBs)) show the presence of both, wild type (*left*) or xlpra1-affected *RPGR* (middle) [partial sequences of exons 4–6 (lane 1) and 3' part of the exon ORF15 (lane 2)] and *ROBO2-AS* (lane 3, partial sequences of exons 2–3). (*right*) RT-PCR results from the negative control experiment where non-specific binding of RNA to MB was tested by same RT-PCR analysis using cDNA generated from RNA attached to MB without 5'-biotin-labelled oligonucleotides. *Note*: The irrelevant lanes were spliced out from the 12-wells agarose gel (indicated by dashed lines).

PCR in selected regions of $RPGR^{orf15}$ and $ROBO2$-AS revealed that lncRNA $ROBO2$-AS was pulled down in complex with both wild type and mutant $RPGR$ mRNAs (Fig. 7B), confirming their targeted interaction.

We previously reported the structure of mouse orthologous $ROBO2$-AS transcript [28]. Here we identified two novel transcriptional variants of human $ROBO2$-AS lncRNA that displayed structural and sequence homology to both, canine $ROBO2$-AS and mouse $Robo2$-AS, and formed a sense-antisense overlap with human $ROBO2$. Both human $ROBO$-AS variants (v. 1 and v. 2) are expressed in the retina (Supplementary Fig. 3A). Comparatively to v. 1, $ROBO2$-AS v. 2 contains additional exon (exon 3, 164 nt) that has high antisense homology (>90%) to human $RPGR$ exon ORF15 (the sequence of exon 3 is shown on Supplementary Fig. 3B). The expression of the $ROBO2$-AS exon 3 was not detected in ARPE19 or Y79 cells. Computational modelling with the IntaRNA software predicted multiple overlapping interaction sites between $ROBO$-AS exon 3 (pos. 450–613 in OP555997) and human ORF15 region (c.1930-c.3250) with the E^{min} varied between −55.9 kcal/mol to −120.8 kcal/mol (Fig. 7C). No RNA-RNA interactions were predicted between $ROBO2$-AS exon 3 and $RPGR$ exons 1–14 (E^{min} is below −24.6 kcal/mol), similar to canine $RPGR^{orf15}$/$ROBO2$-AS pair. Unfortunately, the amount of human retinal RNA was not sufficient for RNA-RNA duplex analysis.

Overall, the enrichment of the potential RNA-RNA interactions between lncRNA $ROBO2$-AS and $RPGR$ exon ORF15 region (in both, canine and human gene pairs), and the presence of the $RPGR$/$ROBO2$-AS RNA/RNA duplex in canine retina suggest the involvement of lncRNA $ROBO2$-AS in regulation of $RPGR$ pre-mRNA processing.

Discussion

To date we still know little about the cellular dysfunction underlying the $RPGR^{orf15}$-dependent XLRP disease. Although it can be a result of disrupted RPGR^{orf15} protein interaction network due to the presence of $RPGR$ mutations [26,53–55], we cannot exclude an impact of impaired regulation of $RPGR^{orf15}$ pre-mRNA processing on disease phenotype. This study has provided a deeper insight into non-canonical splicing events in the $RPGR^{orf15}$ pre-mRNA in canine and human cells. More specifically, our results address the production of novel linear $RPGR$ ORF15-like transcripts and circular $RPGR$ RNAs.

In addition to the retinal $RPGR^{orf15}$ transcript where introns are spliced out through recognition of conserved canonical splice signals (GU, AG) at the 5′ and 3′ ends of the exons, we report several novel ORFs produced from the $RPGR^{orf15}$ pre-mRNA through alternative splicing reactions occurring mostly within ORF14/ORF15 (dog) and ORF15 (human) exonic regions. In these $RPGR$ ORF15-like transcripts portions of the terminal exon sequence were removed through utilization of non-canonical splice sites. Although we found high frequencies of atypical 5' splice site variants, these excised sequences are likely U2-type introns (that are substrates for the major spliceosome) as other $RPGR$ introns. Splice sites are relatively weakly conserved in U2-type introns,

resulting to more flexible splice site choices in alternative splicing processes [56,57].

Contrary to the conclusions of Hong et al. [18] regarding the existence of innumerable $RPGR$ ORF15-like variants, we found a limited number of full-length ORFs corresponding to $RPGR$ ORF15-like transcripts. We deliberately avoided the use of RT-PCR primers designed closely to the $RPGR^{orf15}$ terminal exon to examine splicing heterogeneity within this region (as done in Hong et al., 2002 study [18]), assuming that some of $RPGR$ ORF15-like transcripts might be temporal and be degraded in the nucleus. Although we obtained lesser number of the transcriptional variants our results do not disprove the reported exceptional heterogeneity in $RPGR$ transcripts processing [18]. The presence of $RPGR$ ORF15-like transcripts in the polysome fraction of cultured cells suggests their functionality, as polysomal mRNAs are usually assumed to be actively translated.

Here we found that cis-splicing events in $RPGR$ ORF15 region are accompanied by production of the circRNAs with partially overlapped sequences (in text: circRPGRs). We consider the identified circRPGR sequences to be genuine based on the following reasons: (i) use of a heat-stable reverse transcriptase enzyme with increased thermal stability in the RT reaction (at 55°C, when RT artefacts are usually not detected [58]; (ii) RNaseR treatment; and (iii) the presence of several same circRPGRs in more than one cell type (see Table 1). Therefore, it appears reasonable to conclude that circRPGR transcripts are generated in $vivo$.

The majority of the circRNAs reported to date are generated by back splicing, wherein the 3′ terminus of an exon is ligated to the 5′ terminus of the same or an upstream exon from a single pre-mRNA [33,34,50]. Unlike these 'canonical' circRNAs derived from an exonic back splicing, 'non-canonical' circRNAs (lariat-derived intronic circRNAs, sub-exonic circRNAs, intron circles, etc.) can be produced by non-canonical back fusion events (currently poor defined), for example, through a failure of intron lariat debranching or intramolecular slippage events that require SHSs at the back fusion points for RNA circularization [32,44,59–61]. Sequence analysis of circRPGRs products revealed that none of the identified circRPGRs is a 'canonical' circRNA. Though the exact mechanism underlying the biogenesis of circRPGRs remains unresolved, some of the identified circRPGRs may originate from lariat RNAs excised from the region corresponding to the $RPGR^{orf15}$ terminal exon. In support of this hypothesis, we were able to detect a subset of circRPGRs within the identified exitrons using outward-facing primers (see Fig. 3). Notably, results from our sequence analysis suggest the presence of two different pathways underlying circRPGR biogenesis. Approximately half of circRPGRs have a pair of dinucleotide sequence motifs (different from canonical splice sites) next to their 5′- and 3′-end BFPs. And the other half of circRPGRs contain SHSs adjacent to their BFPs. One may speculate that the presence of two non-canonical pathways (SHS-based and SHS-independent) in circRPGR biogenesis is unusual, however, similar findings were reported in other circRNAs [32]. Moreover, according to previous studies in

eukaryotes, a considerable amount of chimeric linear RNAs have SHSs at the junction sites of the source sequences while other carry dinucleotides at their junctions [41,62,63], suggesting that these processes are probably widespread. Together with canonical *cis*-splicing and back splicing such non-canonical mechanisms contribute to the complexity of the transcriptome.

The presence of circRPGRs in the cytoplasm and polysomal fraction support their functional roles in cellular processes. To explore this possibility, we used a bioinformatic approach to demonstrate their potential functionality. Several m6A consensus motifs were identified in sequences of circRPGRs associated with polysomes. The presence of m6A is reported to promote initiation of protein translation from circRNAs [35,48]. However, whether circRPGRs are translated *in vivo* into protein-like products acting as potential protein decoys to influence cellular functions, still needs to be determined. m6A-driven translation of circRNAs is shown to require activity initiation factor eIF4G2 and m6A reader YTHDF3, and it is enhanced by methyltransferase METTL3/14 [64]. Interestingly, it is also upregulated under stress conditions (such as heat shock) [64]. On the other hand, m6A-containing circRNAs can be subject to endoribonucleolytic cleavage via YTHDF2 (m6A reader protein), HRSP12 (adaptor protein), and RNase P/MRP (endoribonucleases) [65]. These evidences indicate that functionality of circRNAs is under tight regulation and may depend on the physiological state of cells.

In addition, sequence analysis of the *RPGR^orf15* revealed a considerable number of miR-binding sites in the region corresponding *RPGR^orf15* terminal exon. The enrichment of miR-binding sites specifically in the ORF14/ORF15 (dog) and ORF15 (human) suggest that certain circRPGRs generated from this region may act as circular miR-sponges. Similarly, such functional role of circRPGRs requires experimental validation and will be addressed in follow-up studies. Further speculating on this matter, some additional roles for circRPGRs beyond those mentioned above, may be envisioned. One of several possibilities is an involvement of circRPGRs in the regulation of *RPGR^orf15* pre-mRNA processing by competing with the linear transcript for the recognition of specific splicing protein complexes.

The next challenge is to better understand the regulation of the *RPGR^orf15* pre-mRNA that, in fact, is fairly unexplored. To date, there are no publications that document direct interactions between the *RPGR^orf15* terminal exon region and *trans*-acting splicing factors, including RNA-binding proteins and/or non-coding regulatory RNAs. Of those, natural antisense transcripts (NATs) are of particular interest. Among other potential mechanisms of action endogenous NATs can be effectively involved in splicing repression through the formation of RNA–RNA hybrids with target pre-mRNAs, acting in *cis* or in *trans* [66]. In this study we found that in the canine retinal lncRNA *ROBO2-AS* acts in *trans* to form lncRNA/mRNA duplexes with *RPGR*, suggesting that duplex formation could modulate the *RPGR^orf15* pre-mRNA processing. Further validation of these speculations will require genetic studies, including gene knockdown in appropriate settings.

Identifying the spectrum of novel ORF15-like transcripts (linear and circular) as well as a candidate modulator of the *RPGR^orf15* pre-mRNA splicing (lncRNA *ROBO2-AS*) is of considerable clinical interest given that patients with mutations in the *RPGR* exon ORF15 display broad phenotypic variability in disease severity and progression [26,27]. Although challenges are still in place in terms of sequencing difficulties of the ORF15 region, without knowledge about the diversity of ORF15-derived non-canonical splicing products with novel molecular functions, it is difficult to elucidate molecular mechanisms of pathogenic ORF15 mutations. Mutations in the *RPGR* exon ORF15 could not only endow RPGR^orf15 protein with cell-damaging properties but also compromise the ability of circular *RPGR* products originated from the ORF15 genomic locus to perform their functional roles that could itself be a contributing factor into the *RPGR^orf15*-XLRP disease pathogenesis.

It's worth mentioning that since the time when *RPGR^orf15* variant was first reported [5], the expression the *RPGR^orf15* transcript is validated using qRT-PCR primers designed within lesser repetitive part of the exon ORF15 (for example, within 250 nt derived from the extension of exon 15 into intron 15 [16]). However, this ORF15 area is often present in linear *RPGR* ORF15-like transcripts and circRPGRs that are widely expressed. Existence of complex non-canonical events in ORF15 region needs to be taken into consideration to avoid misinterpretation of the experimental data.

Overall, the observations from this study demonstrate complex structural organization of the ORF15 region that multitasks in the biogenesis of multiple transcripts including *RPGR^orf15* as well as other linear and circular ORF15-like transcripts, highlighting the effective mechanisms cells employ for maximizing their genomes. Our findings provide a basis for future mechanistic studies that will inform the molecular events responsible for *RPGR^orf15* pre-mRNA regulation and provide new insights into potential strategies to slow down the disease progression in patients with *RPGR^orf15*-dependent retinal degeneration.

Materials and methods

Ethics statement

The research was conducted in full compliance with the Association for Research in Vision and Ophthalmology (ARVO) Resolution on the Use of Animals in Ophthalmic and Vision Research. All the studies have been approved by the University of Pennsylvania Institutional Animal Care and Use Committee (IACUC).

Study samples

The study used archival fresh frozen retinal samples (16–24 weeks old) from normal dogs and xlpra1&2 carriers mostly remaining from the previously published studies [67–69]. Normal canine brain tissue (1.2 years old) was collected and stored at −80°C. Fresh retinal tissue from phenotypically normal dog (6 months old) from the study supported by the

NIH grant P40-OD-010939-36 (Vite PI) was kindly provided by Dr. Charles Vite (University of Pennsylvania). Canine primary skin fibroblasts were isolated from skin punch biopsy (3 mm) from normal and xlpra2 mutation affected dogs and cultured as described below. OCT embedded normal canine retinal sections used for immunohistochemistry were samples from current previous lab projects. Retinal RNA was isolated from the human eye (67 years old, male) acquired from the Rochester Eye & Tissue Bank (Rochester, NY, USA) and the archival human retinal RNA sample left from previous projects was used in this study. Human cell lines Y79 and ARPE19 were obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA) and were cultured in recommended growth conditions. The aliquots from Y79 and ARPE19 cells were a kind gift of Drs. Raghavi Sudharsan and Karina Guziewicz, respectively, both scientists are members of our lab group and had purchased the cells directly from ATCC.

Isolation and culture of canine primary skin fibroblasts

Fibroblasts were obtained by digesting the skin punch biopsy samples in dispase and collagenase IV (both from STEMCELL Technologies Inc., Cambridge, MA, USA) enzymes mixture (2.5 hours at 37°C) and undigested tissues were filtered through a 100 μm cell strainer. The cell suspensions were diluted in DMEM/F12 medium (ATCC) and centrifuged at 300 g for 8 min. Collected cells were grown in DMEM/F12 medium supplemented with 10% foetal bovine serum (FBS) (Sigma-Aldrich, Saint-Louis, MO, USA), penicillin/streptomycin and amphotericin B (Sigma-Aldrich) at 37°C in a humidified 5% CO_2 incubator.

Subcellular fractionation, RNA extraction and cDNA synthesis

Cytoplasmic fractions (from canine primary skin fibroblasts, Y79 and ARPE19 cells) were isolated using a PARIS kit (Thermo Fisher Scientific, Waltham, MA, USA). Polysomal fractions (from the collected fresh retina (6 months old), canine primary skin fibroblasts, Y79 and ARPE19 cells) were obtained according to previously published protocol [70] using 1 M sucrose cushion. Polysomes were pelleted by centrifugation in a TLA100.3 rotor at 50,000 rpm, 4°C for 4 hours. RNA from cells or subcellular fractions (cytoplasm, polysomes) was isolated using a modified TRIzol and single chloroform extraction protocol as previously described [71]. First strand cDNA was synthesized using the Maxima H Minus cDNA Synthesis Kit (Thermo Fisher Scientific) following the manufacturer's recommendations. Reverse transcription (RT) was performed at 55°C.

RNase R treatment

2 μg total RNA was incubated at 37°C (1–2.5 hours) or 45°C (2 hours) in 20 μl reactions that contained 2 μl 10× RNase R Buffer (0.2 M Tris–HCl (pH 8.0), 1 mM MgCl2 and 1 M KCl or LiCl), 10 U RiboLock RNase Inhibitor (Thermo Fisher Scientific) and 2 U RNase R (Lucigen, Middleton,

WI, USA). Reactions were purified with RNA Clean & Concentrator-5 (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions and the RNA was eluted in 15 μl nuclease-free water. *Note*: The increased temperature (45°C) was recommended by the technical support team (Lucigen).

Reverse transcription Polymerase Chain Reaction (RT-PCR)

RT-PCR was performed using GoTaq® Long PCR Master Mix (Promega, Madison, WI, USA) using cDNA generated from 25–50 ng of RNA and a primer concentration of 0.8 μM. Corresponding primers sequences are listed in Supplementary Table S5. Betaine solution (Sigma-Aldrich) was added to all reaction to final concentration 1.5 M. PCR thermal cycler conditions were: (1) denaturation 30 sec at 95°C, annealing 40 sec at 64°C and elongation 1 min or 5 min at 68°C (for *ROBO2-AS* and *RPGR* RT-PCR products, respectively) and (2) denaturation 30 sec at 95°C, annealing 45 sec at 60°C and elongation 1 min at 68°C (for circular RNAs). PCR steps were repeated 37–39 times following by agarose gel purification and direct sequencing of PCR products. *Note*: For Sanger sequencing of long-range PCR products a set of internal primers was used for directed sequencing of entire PCR products. PCR based cloning was generally avoided and only used twice to clarify partially noisy sequencing reactions.

Western blot analysis

Western blots were carried out as previously described [67] using N-terminal anti-RPGR antibodies, (Abclonal, Woburn, MA, USA) at 1:500 ratio. Protein concentrations were determined by BCA Protein Assay (Thermo Fisher Scientific), and equal micrograms of protein analysed. Quantification of proteins on western blot was performed using Image Studio Software provided by LI-COR. Western blot was done in total retinal protein extracts in normal (16 weeks old) and xlpra1 carrier (20 weeks) retinas.

Fluorescent immunohistochemistry (IHC)

The procedure was carried out as previously described [72]. Briefly, 10 μm retinal cryosections from PFA fixed and 20 μm retinal cryosections from unfixed tissues were washed and treated with primary antibodies (anti-RPGR (Abclonal) at 1:200 ratio and anti-polyglutamylation modification GT335 (Adipogen Life Sciences, Inc., San Diego, CA, USA) at 1:1000 ratio) in PBS solution, 3% normal horse serum, 1% BSA and 0.3% Triton X-100 overnight followed by incubation with appropriate fluorescent secondary antibodies (Alexa Fluor Dyes, 1:200; Molecular Probes, Eugene, OR, USA). Confocal images were captured with TCS-SP5 confocal microscope system (Leica Microsystems, Buffalo Grove, IL, USA) under identical conditions and imported into ImageJ software [73]. Confocal images are shown as a z-stack of 11 z-steps or as a single z-step each of 0.21 μm. Maximum projection of all

images was equally adjusted for contrast and brightness with ImageJ software.

RNA pull-down assay using magnetic beads

200 μg of streptavidin magnetic beads S1421 (New England Biolabs, Ipswich, MA, USA) were resuspended 100 μl of Wash/Binding Buffer 1 (WBB1) [0.5 M NaCl, 20 mM Tris-HCl (pH 7.5), 1 mM EDTA]. 5' biotinylated canine *RPGR*-specific oligonucleotides (listed in Supplementary Table S5) were dissolved in 25 μl of WBB1 to final concentration 30 μM (each), added to magnetic beads and incubated at room temperature for 5 min with occasional agitation. Beads were washed twice by adding 100 μl of WBB1 and finally dissolved in WBB2 [150 mM NaCl, 20 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM DTT and RNase inhibitor (New England Biolabs)]. 6 μg of total RNA (isolated from 24 weeks old normal canine retina) was added to the beads and incubated at room temperature 3 min. The beads/oligonucleotides/RNA complex was washed in WBB2 for 10 min with occasional agitation by hand, following additional six washes, 5 min each. After applying magnet and removal of supernatant, beads were re-suspended in 25 μl of water. The suspension of beads with RNA hybridized was used in full in subsequent RT reaction followed by 34 cycles of RT-PCR.

Databases and sequence analysis tools

Leiden Open Variation Database (LOVD) [https://databases.lovd.nl/shared/genes/RPGR] and National Library of Medicine dbSNP [https://www.ncbi.nlm.nih.gov/snp/] were used to obtained sequence variants in human *RPGR*. miRDB online software [74] was used for miRNA targets prediction [http://mirdb.org/custom.html]. The IntaRNA web-based software [75] was used for the prediction of RNA-RNA interactions [http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp]. DNASTAR Lasergene package (v. 13), the UCSC Genome Browser [76] [https://genome.ucsc.edu/] and the UniProt [77] [uniprot.org/] were used for detailed sequence analysis and alignments.

Data availability

All novel linear transcripts data generated in this study have been deposited to the NCBI GenBank database and Accession Numbers are listed in Supplementary Table S1. Circular RNA data are available within the article or its supplementary materials.

ORCID

Tatyana Appelbaum http://orcid.org/0000-0001-8765-4792

References

[1] Sahel J, Bonnel S, Mrejen S, et al. Retinitis pigmentosa and other dystrophies. Dev Ophthalmol. 2010;47:160–167.

[2] Breuer DK, Yashar BM, Filippova E, et al. comprehensive mutation analysis of RP2 and RPGR in a North American cohort of families with X-linked retinitis pigmentosa. Am J Hum Genet. 2002;70(6):1545–1554.

[3] Ott J, Bhattacharya S, Chen JD, et al. Localizing multiple X chromosome-linked retinitis pigmentosa loci using multilocus homogeneity tests. Proc Natl Acad Sci U S A. 1990;87(2):701–704.

[4] Meindl A, Dry K, Herrmann K, et al. A gene (RPGR) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X-linked retinitis pigmentosa (RP3). Nat Genet. 1996;13(1):35–42.

[5] Vervoort R, Lennon A, Bird AC, et al. Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. Nat Genet. 2000;25(4):462–466.

[6] Shu X, McDowall E, Brown AF, et al. The human retinitis pigmentosa GTPase regulator gene variant database. Hum Mutat. 2008;29(5):605–608.

[7] Churchill JD, Bowne SJ, Sullivan LS, et al. Mutations in the X-linked retinitis pigmentosa genes RPGR and RP2 Found in 8.5% of families with a provisional diagnosis of autosomal dominant retinitis pigmentosa. Invest Ophthalmol Vis Sci. 2013;54(2):1411–1416.

[8] Nguyen XT, Talib M, van Schooneveld MJ, et al. RPGR-associated dystrophies: clinical, genetic, and histopathological features. Int J Mol Sci. 2020;21(3):835.

[9] Ayyagari R, Demirci FY, Liu J, et al. X-linked recessive atrophic macular degeneration from RPGR mutation. Genomics. 2002;80(2):166–171.

[10] Yang Z, Peachey NS, Moshfeghi DM, et al. Mutations in the RPGR gene cause X-linked cone dystrophy. Hum Mol Genet. 2002;11(5):605–611.

[11] Beltran WA, Cideciyan AV, Boye SE, et al. Optimization of retinal gene therapy for X-linked retinitis pigmentosa due to RPGR mutations. Mol Ther. 2017;25(8):1866–1880.

[12] Pawlyk BS, Bulgakov OV, Sun X, et al. Photoreceptor rescue by an abbreviated human RPGR gene in a murine model of X-linked retinitis pigmentosa. Gene Ther. 2016;23(2):196–204.

[13] Ferreira PA. Insights into X-linked retinitis pigmentosa type 3, allied diseases and underlying pathomechanisms. Hum Mol Genet. 2005;14(2):R259–67. Spec No.

[14] Wright AF, Shu X. Focus on molecules: RPGR. Exp Eye Res. 2007;85(1):1–2.

[15] Wright RN, Hong DH, Perkins B. Misexpression of the constitutive Rpgr ex1-19 variant leads to severe photoreceptor degeneration. Invest Ophthalmol Vis Sci. 2011;52(8):5189–5201.

[16] Moreno-Leon L, West EL, O'Hara-Wright M, et al. RPGR isoform imbalance causes ciliary defects due to exon ORF15 mutations in X-linked retinitis pigmentosa (XLRP). Hum Mol Genet. 2021;29(22):3706–3716.

[17] Raghupathy RK, Gautier P, Soares DC, et al. Evolutionary characterization of the retinitis pigmentosa GTPase regulator gene. Invest Ophthalmol Vis Sci. 2015;56(11):6255–6264.

[18] Hong DH, Li T. Complex expression pattern of RPGR reveals a role for purine-rich exonic splicing enhancers. Invest Ophthalmol Vis Sci. 2002;43(11):3373–3382.

[19] Pawlyk BS, Bulgakov OV, Liu X, et al. Replacement gene therapy with a human RPGRIP1 sequence slows photoreceptor degeneration in a murine model of leber congenital amaurosis. Hum Gene Ther. 2010;21(8):993–1004.

[20] Hong DH, Pawlyk B, Sokolov M, et al. RPGR isoforms in photoreceptor connecting cilia and the transitional zone of motile cilia. Invest Ophthalmol Vis Sci. 2003;44(6):2413–2421.

[21] Vervoort R, Wright AF. Mutations of RPGR in X-linked retinitis pigmentosa (RP3). Hum Mutat. 2002;19(5):486–500.

[22] Zhang Q, Acland GM, Wu WX, et al. Different RPGR exon ORF15 mutations in canids provide insights into photoreceptor cell degeneration. Hum Mol Genet. 2002;11(9):993–1003.

[23] Thompson DA, Khan NW, Othman MI, et al. Rd9 is a naturally occurring mouse model of a common form of retinitis pigmentosa caused by mutations in RPGR-ORF15. PLoS One. 2012;7(5): e35865.

[24] Huang XF, Wu J, Lv JN, et al. Identification of false-negative mutations missed by next-generation sequencing in retinitis pigmentosa patients: a complementary approach to clinical genetic diagnostic testing. Genet Med. 2015;17(4):307–311.

[25] De La Camara C M-F, Nanda A, Salvetti AP, et al. Gene therapy for the treatment of X-linked retinitis pigmentosa. Expert Opin Orphan Drugs. 2018;6(3):167–177.

[26] Fahim AT, Bowne SJ, Sullivan LS, et al. Allelic heterogeneity and genetic modifier loci contribute to clinical variation in males with X-linked retinitis pigmentosa due to RPGR mutations. PLoS One. 2011;6(8):e23021.

[27] Charng J, Cideciyan AV, Jacobson SG, et al. Variegated yet non-random rod and cone photoreceptor disease patterns in RPGR-ORF15 -associated retinal degeneration. Hum Mol Genet. 2016;25(24):5444–5459.

[28] Appelbaum T, Murgiano L, Becker D, et al. Candidate genetic modifiers for RPGR retinal degeneration. Invest Ophthalmol Vis Sci. 2020;61(14):20.

[29] Vaquero-Garcia J, Barrera A, Gazzara MR, et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. Elife. 2016;5:e11752.

[30] Buskin A, Zhu L, Chichagova V, et al. Disrupted alternative splicing for genes implicated in splicing and ciliogenesis causes PRPF31 retinitis pigmentosa. Nat Commun. 2018;9(1). DOI:10.1038/s41467-018-06448-y.

[31] Liang D, Wilusz JE. Short intronic repeat sequences facilitate circular RNA production. Genes Dev. 2014;28(20):2233–2247.

[32] Liu X, Hu Z, Zhou J, et al. Interior circular RNA. RNA Biol. 2020;17(1):87–97.

[33] Barrett SP, Salzman J. Circular RNAs: analysis, expression and potential functions. Development. 2016;143(11):1838–1847.

[34] Xiao M-S, Ai Y, Wilusz JE. Biogenesis and functions of circular RNAs come into focus. Trends Cell Biol. 2020;30(3):226–240.

[35] Yang Y, Fan X, Mao M, et al. Extensive translation of circular RNAs driven by N(6)-methyladenosine. Cell Res. 2017;27(5):626–641.

[36] Hanineva A, Park KS, Wang JJ, et al. Emerging roles of circular RNAs in retinal diseases. Neural Regen Res. 2022;17(9):1875–1880.

[37] Wright RN, Hong DH, Perkins B. Rpgr ORF15 connects to the usher protein network through direct interactions with multiple whirlin isoforms. Invest Ophthalmol Vis Sci. 2012;53(3):1519–1529.

[38] Shi H, Zhou Y, Jia E, et al. Comparative analysis of circular RNA enrichment methods. RNA Biol. 2022;19(1):55–67.

[39] Xiao MS, Wilusz JE. An improved method for circular RNA purification using RNase R that efficiently removes linear RNAs containing G-quadruplexes or structured 3' ends. Nucleic Acids Res. 2019;47(16):8755–8769.

[40] Eger N, Schoppe L, Schuster S, et al. Circular RNA splicing. Adv Exp Med Biol. 2018;1087:41–52.

[41] Li X, Zhao L, Jiang H, et al. Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes. J Mol Evol. 2009;68(1):56–65.

[42] Yang W, Wu JM, Bi AD, et al. Possible formation of mitochondrial-RNA containing chimeric or trimeric RNA implies a post-transcriptional and post-splicing mechanism for RNA fusion. PLoS One. 2013;8(10):e77016.

[43] Robic A, Demars J, Kühn C. In-depth analysis reveals production of circular RNAs from non-coding sequences. Cells. 2020;9(8):1806.

[44] Robic A, Kühn C. Beyond back splicing, a still poorly explored world: non-canonical circular RNAs. Genes (Basel). 2020;11(9):1111.

[45] Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. J Appl Genet. 2018;59(3):253–268.

[46] Ashwal-Fluss R, Meyer M, Pamudurti NR, et al. circRNA biogenesis competes with pre-mRNA splicing. Mol Cell. 2014;56(1):55–66.

[47] Barrett SP, Wang PL, Salzman J. Circular RNA biogenesis can proceed through an exon-containing lariat precursor. Elife. 2015;4:e07540.

[48] Pamudurti NR, Bartok O, Jens M, et al. Translation of CircRNAs. Mol Cell. 2017;66(1):9–21.e7.

[49] Abe N, Matsumoto K, Nishihara M, et al. Rolling circle translation of circular RNA in living human cells. Sci Rep. 2015;5(1):16435.

[50] Yesharim L, Mojbafan M, Abiri M. Hints from the cellular functions to the practical outlook of circular RNAs. Front Genet. 2021;12:679446.

[51] Auweter SD, Oberstrass FC, Allain FH. Sequence-specific binding of single-stranded RNA: is there a code for recognition? Nucleic Acids Res. 2006;34(17):4943–4959.

[52] Saberi F, Kamali M, Najafi A, et al. Natural antisense RNAs as mRNA regulatory elements in bacteria: a review on function and applications. Cell Mol Biol Lett. 2016;21(1):6–016-0007–z. eCollection 2016.

[53] Patnaik SR, Raghupathy RK, Zhang X, et al. The role of RPGR and its interacting proteins in ciliopathies. J Ophthalmol. 2015;2015:414781.

[54] Murga-Zamalloa CA, Swaroop A, Khanna H. RPGR-containing protein complexes in syndromic and non-syndromic retinal degeneration due to ciliary dysfunction. J Genet. 2009;88(4):399–407.

[55] Zhang Q, Giacalone JC, Searby C, et al. Disruption of RPGR protein interaction network is the common feature of RPGR missense variations that cause XLRP. Proc Natl Acad Sci U S A. 2019;116(4):1353–1360.

[56] Sibley CR, Blazquez L, Ule J. Lessons from non-canonical splicing. Nat Rev Genet. 2016;17(7):407–421.

[57] Akinyi MV, Frilander MJ. At the intersection of major and minor spliceosomes: crosstalk mechanisms and their impact on gene expression. Front Genet. 2021;12:700744.

[58] Cocquet J, Chong A, Zhang G, et al. Reverse transcriptase template switching and false alternative transcripts. Genomics. 2006;88(1):127–131.

[59] Zhang Y, Zhang XO, Chen T, et al. Circular intronic long non-coding RNAs. Mol Cell. 2013;51(6):792–806.

[60] Robic A, Cerutti C, Kühn C, et al. Comparative analysis of the circular transcriptome in muscle, liver, and testis in three livestock species. Front Genet. 2021;12:665153.

[61] Taggart AJ, Lin CL, Shrestha B, et al. Large-scale analysis of branchpoint usage across species and cell lines. Genome Res. 2017;27(4):639–649.

[62] Ritz K, van Schaik BD, Jakobs ME, et al. Looking ultra deep: short identical sequences and transcriptional slippage. Genomics. 2011;98(2):90–95.

[63] Dubrovina AS, Kiselev KV, Zhuravlev YN. The role of canonical and noncanonical pre-mRNA splicing in plant stress responses. Biomed Res Int. 2013;2013:264314.

[64] Yang T, Huang H, Shao Q, et al. miR-92 suppresses Robo1 translation to modulate slit sensitivity in commissural axon guidance. Cell Rep. 2018;24(10):2694–2708.e6.

[65] Park OH, Ha H, Lee Y, et al. Endoribonucleolytic cleavage of m(6) A-containing RNAs by RNase P/MRP complex. Mol Cell. 2019;74 (3):494–507.e8.

[66] Romero-Barrios N, Legascue MF, Benhamed M, et al. Splicing regulation by long noncoding RNAs. Nucleic Acids Res. 2018;46 (5):2169–2184.

[67] Appelbaum T, Becker D, Santana E, et al. Molecular studies of phenotype variation in canine RPGR-XLPRA1. Mol Vis. 2016;22:319–331.

[68] Genini S, Guziewicz KE, Beltran WA, et al. Altered miRNA expression in canine retinas during normal development and in models of retinal degeneration. BMC Genomics. 2014;15 (1):172–2164-15-172.

[69] Genini S, Beltran WA, Aguirre GD. Up-regulation of tumor necrosis factor superfamily genes in early phases of photoreceptor degeneration. PLoS One. 2013;8(12):e85408.

[70] Ingolia NT, Brar GA, Rouskin S, et al. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc. 2012;7 (8):1534–1550.

[71] Appelbaum T, Santana E, Aguirre GD. Strong upregulation of inflammatory genes accompanies photoreceptor demise in canine models of retinal degeneration. PLoS One. 2017;12(5):e0177224.

[72] Beltran WA, Hammond P, Acland GM, et al. A frameshift mutation in RPGR exon ORF15 causes photoreceptor degeneration and inner retina remodeling in a Model of X-Linked retinitis pigmentosa. Invest Ophthalmol Vis Sci. 2006;47(4):1669–1681.

[73] Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671–675.

[74] Chen Y, Wang X. Wang X. miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res. 2020;48(D1):D127–D131.

[75] Mann M, Wright PR, Backofen R. IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. Nucleic Acids Res. 2017;45(W1):W435–W439.

[76] Navarro Gonzalez J, Zweig AS, Speir ML, et al. The UCSC genome browser database: 2021 update. Nucleic Acids Res. 2021;49(D1):D1046–D1057.

[77] Consortium U. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49(D1):D480–D489.