

Article

# Volatile Evolution of Long Non-Coding RNA Repertoire in Retinal Pigment Epithelium: Insights from Comparison of Bovine and Human RNA Expression Profiles

Olga A. Postnikova <sup>1</sup>, Igor B. Rogozin <sup>2</sup> , William Samuel <sup>1</sup>, German Nudelman <sup>3</sup>, Vladimir N. Babenko <sup>4</sup>, Eugenia Poliakov <sup>1</sup> and T. Michael Redmond <sup>1,\*</sup> 

<sup>1</sup> Laboratory of Retinal Cell and Molecular Biology, National Eye Institute, National Institutes of Health, Bethesda, MD 20892, USA; olga.postnikova@nih.gov (O.A.P.); Samuelw@nei.nih.gov (W.S.); poliakove@nei.nih.gov (E.P.)

<sup>2</sup> National Center for Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD 20894, USA; rogozin@ncbi.nlm.nih.gov

<sup>3</sup> Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; german.nudelman@mssm.edu

<sup>4</sup> Laboratory of Neuropathology Modeling, The Federal Research Center Institute of Cytology and Genetics SB RAS, Novosibirsk 630090, Russia; bob@bionet.nsc.ru

\* Correspondence: redmondd@helix.nih.gov

Received: 31 January 2019; Accepted: 26 February 2019; Published: 8 March 2019



**Abstract:** Currently, several long non-coding RNAs (lncRNAs) (TUG1, MALAT1, MEG3 and others) have been discovered to regulate normal visual function and may potentially contribute to dysfunction of the retina. We decided to extend these analyses of lncRNA genes to the retinal pigment epithelium (RPE) to determine whether there is conservation of RPE-expressed lncRNA between human and bovine genomes. We reconstructed bovine RPE lncRNAs based on genome-guided assembly. Next, we predicted homologous human transcripts based on whole genome alignment. We found a small set of conserved lncRNAs that could be involved in signature RPE functions that are conserved across mammals. However, the fraction of conserved lncRNAs in the overall pool of lncRNA found in RPE appeared to be very small (less than 5%), perhaps reflecting a fast and flexible adaptation of the mammalian eye to various environmental conditions.

**Keywords:** lncRNA; RPE; evolutionary conservation; bovine lncRNA

## 1. Introduction

Long non-coding RNAs (lncRNA) are transcripts that do not encode proteins and are longer than 200 nucleotides. According to GENCODEv7, the human genome has 9277 manually annotated lncRNA genes producing 14,880 transcripts [1]. lncRNAs are transcribed by polymerase II and can be polyadenylated and spliced, just as protein coding transcripts. It has been demonstrated that lncRNAs are more tissue-specific and are expressed, on average, approximately 10 times less compared to protein coding mRNAs [2]. While some lncRNAs have described regulatory roles most are functionally uncharacterized [3–9]. The idea that most lncRNAs are by-products of background transcription, and “simply the noise emitted by a busy machine” [10,11], may be due to the low abundance and poor evolutionary conservation of many lncRNAs compared to protein-coding sequences and small RNAs (e.g., miRNAs and snoRNAs) [12]. However, some lncRNAs have strongly conserved regions [13]. Also, most lncRNAs show reduced rates of substitution and insertion/deletion that suggest purifying selection [6,14,15].

As strong sequence conservation is lacking, identification of lncRNAs at the genome level depends on expression analysis. This makes a complete annotation of the mammalian lincRNome difficult to achieve. By using a variety of approaches, there has been ongoing identification of novel lncRNAs from the transcriptomes of several mammalian taxa [16]. The FANTOM3 project, in particular, has identified >30,000 putative lncRNAs in mouse tissues by full-length cDNA cloning [17,18]. However, an important unsolved issue in these analyses is a lack of a clear boundary between protein-coding and true lncRNA genes. Using the support vector machine technique, the FANTOM3 project transcripts have been separated into coding and non-coding transcripts and thereby applied to estimate how many lncRNAs occur in a mouse. Accordingly, ~14,000 lnc RNAs were newly identified in the FANTOM3 data while the total number of lncRNAs was estimated to be ~28,000 [19]. However, Ji and colleagues showed that many cytoplasmic lncRNA are bound to ribosomes and may be translated to small peptides [20]. Thus, the question of the protein-coding potential of lncRNAs remains open.

Managadze et al. compared the identified human and mouse long intergenic non-coding (linc)RNAs using a new statistical technique to estimate their size and evolutionary conservation [21]. They estimate that human and mouse lincRNAs outnumber protein-coding genes by at least two to one. Furthermore, there is apparent conservation of two-thirds of the lincRNAs between human and mouse, although criteria of orthology were relatively relaxed in this study (100 nucleotide overlap of putative orthologs of human lincRNAs with the mouse lincRNAs and liberal expression cut-off values [21]). This suggests that there exists, as yet, thousands of conserved but uncharacterized functions [21] and is consistent with several other studies that used more stringent criteria of lncRNA conservation [14,22–25]. Furthermore, analysis of intron-exon structure in lincRNA genes from various mammalian genomes suggests the conservation of introns for over 100 million years [26]. This result suggested that the primary and/or secondary structure of these molecules may be functionally important [26]. Other evolutionary studies suggested that the fraction of conserved lncRNA genes (shared by primates and other mammals) is slightly larger than 20% [23,27], thus various approaches still tend to produce large quantities of evolutionary conserved lncRNA genes in mammalian genomes.

However, conservation of lncRNAs and their gene structure appeared to be tissue-specific. For example, only 18 retinal lncRNAs were found highly conserved between human and mice [28]. Cell type-specific lncRNAs also were shown to be expressed following differentiation of human induced pluripotent stem cells (iPS) into retinal pigment epithelium (RPE) [29]. Currently there are many examples of lncRNAs (TUG1, MALAT1, MEG3 and others) that regulate proper visual function and that may contribute to retinal dysfunction [30]. We decided to extend these analyses of lncRNA genes to the retinal pigment epithelium (RPE). RPE is a monolayer of pigmented cells located between the choroid and retina of the eye. RPE cells are crucial for absorption of the light, transport of nutrients and ions, retinoid visual cycle and phagocytosis of photoreceptor outer segments. These very specific functions of RPE in the eye determine its unique gene expression. We were interested to uncover conserved lncRNAs that could be involved in signature RPE functions and evaluate the fraction of such lncRNAs in the overall pool of lncRNAs found in the RPE. Human eye lncRNAs are not very well characterized, partially because of difficulties with tissue accessibility. Additionally, and compounding this difficulty, human eye layers (choroid, RPE and retina) are notoriously difficult to separate from each other.

Both retina and RPE are derived from neuroectoderm, and they co-differentiate and influence each other. This process must be tightly regulated. Accordingly, lncRNAs are known to be involved in retinal cell fate determination [31], and the development of retinal architecture [32]. In this study, because many eye lncRNAs have tissue-specific expression [33], and it is very difficult to obtain pure human native RPE from donor eyes, we studied predicted lncRNA genes in bovine RPE to understand their functional roles and evolutionary conservation. To predict lncRNAs, we used a tool called FEELnc, which also classifies them with respect to their distance and strand orientation to the closest protein-coding genes. We found a very small set of conserved lncRNAs in the overall pool of lncRNA

that could be involved in signature RPE functions conserved across mammals. This may reflect a fast and flexible adaptation of the mammalian eye to various environmental conditions.

## 2. Materials and Methods

### 2.1. Datasets and Cell Cultures

RNA-seq data used in this study are described in Table 1. Cultured human ARPE-19 cells were grown and differentiated as previously described in Samuel et al. [34].

**Table 1.** RNA-seq datasets.

Data Set	Reads/Library	Description	Publication
Nasal, temporal, and macular regions of human retinal pigment epithelium (RPE)/choroid	2 × 100 bp Paired-end stranded	Retina and RPE/choroid from the temporal, macular, and nasal regions of four clinically normal human donor eyes	Whitmore et al. 2014 [35]
Human ARPE-19 cells	2 × 50 bp Paired end rRNA depleted	ARPE-19 cells cultured for 4 months	Samuel et al. 2017 [34]
Bovine RPE cells	2 × 50 bp Paired end rRNA depleted	Fresh native bovine RPE; bovine RPE cultured for 4 and 8 weeks	This study

### 2.2. Reconstruction and Analysis of Expression of Novel Bovine Long Non-Coding RNAs (lncRNAs)

First reads from native and cultured bovine RPE were aligned with Star 2.5 (default parameters were used; [36]) to UMD3.1 reference bovine genome. These alignments were submitted to Stringtie (default parameters were used; [37]) to obtain reference-based assembled transcripts for each sample. Consequently, transcripts missing from the reference genome were merged across samples to create a novel putative set of lncRNAs for further classification with FEELnc (default parameters were used; [38]). This step filters all protein coding transcripts, computes the coding potential of candidate transcripts and classifies lncRNAs based on their genomic localization. Also, the statistical model for the coding potential was trained on the bovine NONCODE data (23,696 bovine lncRNAs). All transcripts assembled from single exons were filtered out prior to the analysis. We also filtered out transcripts that were shorter than 200 bp according to the commonly accepted definition of long noncoding RNAs.

### 2.3. Reconstruction of Putative Orthologous Regions in the Human Genome

The genomic coordinates and sequences of exons and introns of reconstructed bovine lncRNA genes were obtained using the BLAT program (<https://genome.ucsc.edu/FAQ/FAQblat>) applied to the bovine bosTau8 assembly downloaded from the University of California, Santa Cruz, UCSC Table Browser (<http://hgdownload.soe.ucsc.edu/downloads.html>). Pairwise alignments of these regions in bovine (bosTau8) and human (hg19) genomes were fetched using the Galaxy system [39]. Sequences of putative human orthologous lncRNA exons were extracted from these genomic alignments using an ad hoc program. The genomic coordinates of putative human orthologous lncRNA exons/genes were obtained using the BLAT program (<https://genome.ucsc.edu/FAQ/FAQblat>) applied to the human hg19 genome assembly downloaded from the UCSC Table Browser (<http://hgdownload.soe.ucsc.edu/downloads.html>). It should be noted that the prediction of putative human orthologous lncRNA exons/genes depends on the quality of genomic alignments. Unfortunately, there are well known problems with the accuracy of genomic alignments [40] thus we expect that at least some putative human orthologous lncRNA exons/genes will be missed or incomplete.

To control for the possibility of contamination of the lncRNA data set with previously unidentified protein-coding genes, an additional test was performed: the coding potential of lncRNAs was predicted

using the SYNCODE program using pairwise alignments of bovine and human lncRNA genes [15,41]. All lncRNAs showing signs of evolutionary conservation similar to protein-coding genes were removed from the lncRNA set following the strategy suggested in previous studies [2,27].

#### 2.4. Analysis of Expression of Putative Human lncRNA Genes

A new GTF file for the human genome was created after adding novel putative lncRNAs. We used dbgap RNA-seq data from human RPE/choroid and retina to elucidate how many predicted lncRNA were expressed in native human RPE. Reads were mapped onto the human GRCh38.9 reference genome and counted with “CLC Work Bench”. Differentially expressed transcripts/genes (DEGs) were predicted with DESeq2 (default parameters were used; [42]). Transcripts were considered to be DEGs when fold change >2 and FDR < 0.01.

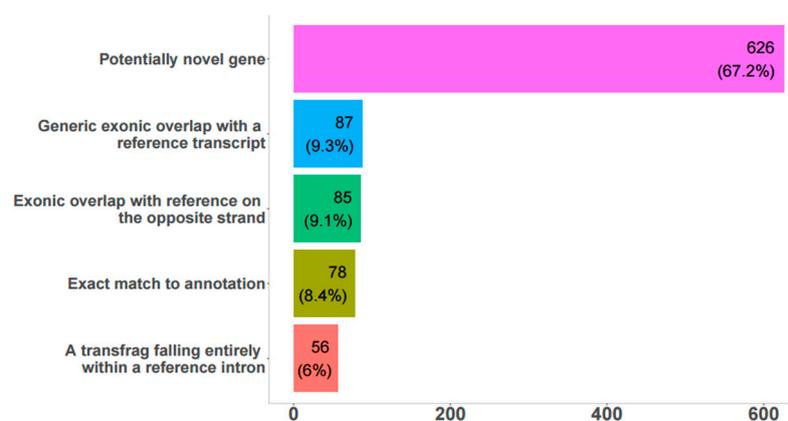
#### 2.5. Confirmation of lncRNA Transcript Sequences

RNA was extracted from bovine RPE and ARPE-19 cultured human RPE cells as previously described. Two sets of cDNA were prepared separately using Quantum oligo dT primer or random hexamers. For all samples no-RT controls were generated and used for polymerase chain reaction (PCR) with all lncRNAs. Since all tested lncRNAs appeared to be polyadenylated, only polyA cDNA was used for PCR. When possible, primers were designed to the 3′- or 5′-most ends of predicted transcripts. Also, short read sequencing was used to design primers to check for fusion presence, lncRNA/untranslated region (UTR), and protein coding genes. PCR products were subjected to gel purification if multiple bands were present. All PCR products were TA cloned in TOPO vector and sequenced. Transcript start and stop sites of human *NEAT1* (*MSTRG.13090.1* and *MSTRG.13090.2*) transcripts expressed in ARPE-19 cells were determined using rapid amplification of cDNA ends (RACE; Clontech) technique according to the manufacturer’s protocol. Novel full length transcript was submitted to Gen Bank with ID MK562403.

### 3. Results

#### 3.1. Reconstruction and Analysis of Expression of Novel Bovine lncRNAs

The total number of transcripts predicted from FEELnc was 848 (656 genes), all of which were potential bovine RPE lncRNAs (Figure 1, Supplementary File S1.GTF).



**Figure 1.** Comparison of detected long non-coding RNAs (lncRNA) to reference genome annotation. Bar graph showing number and percentage of transcripts obtained from FEELnc analysis with low coding potential that belongs to each annotation according to the UMD3.1 reference bovine genome.

Among these, about 67% were potentially novel genes, 8% were an exact match to annotated genes, and another 25% were partially overlapping with transcripts in the reference genomes on the sense or antisense strand. These 848 lncRNAs were classified as genic (157), intergenic (542), and lncRNAs spanning both annotated genes and extending into intergenic space (149), (Supplementary Table S1). 67% (572) of lncRNAs were expressed in native bovine RPE, according to RNA-seq data. In general, there was very good agreement between predicted transcripts in the bovine genome and sashimi plots in IGV (Supplementary Figure S1); 106 and 159 unique differentially expressed (DE) lncRNAs were found between fresh and 4- and 8-week cultured RPE, respectively; 233 DE lncRNAs were common between both time points. Only 23 transcripts had significant hits with the RFAM database (Supplementary Table S2), and some of these have known important biological roles, including MALAT1, TUG1, MEG3, H19, Xist, Six3os1, and NEAT1.

### 3.2. Reconstruction and Analysis of Expression of Putative Orthologous Human lncRNA Genes

In our analysis, 642 gene regions in the human genome with substantial homologies ( $\geq 60\%$  identity) were annotated as corresponding to bovine lncRNAs. It was predicted that 818 transcripts potentially could be transcribed from these genes (Supplementary File S2.GTF). There were 698 transcripts longer than 200 bp, with the longest one 8243 bp. We compared this data set with novel lncRNAs described in fetal RPE and induced Pluripotent Stem Cells (iPSCs) [29] and did not find any overlap. This could be explained by the different focuses of the studies. The set of lncRNAs from fetal or iPSC-RPE excludes all known annotated lncRNA. Moreover, adult human native RPE has its own novel set of lncRNA, that are not expressed in fetal or iPSC derived RPE cells.

To determine expression level of lncRNAs in native RPE and retina we used the dataset from Whitmore et al. 2014 [35]. There were 38 regions homologous to bovine lncRNA genes with an average of 20 reads mapped for all RPE samples. These 38 genes we considered to be expressed in native RPE with higher probability. Also, 54 genes were differentially expressed between the human retina and RPE samples. Because we were interested in determining whether lncRNAs are involved in differentiation and maintenance of RPE, we used our bovine data to identify possible candidates. We found 37 potential lncRNAs to be differentially expressed lncRNAs both between RPE and the retina, and between native bovine RPE and cultured bovine RPE cells. To this list we added 12 lncRNAs that were expressed in native RPE and retina at the same level but whose expression level changed when RPE was explanted and cultured. In this way, a list of 49 potentially functional RPE lncRNAs was obtained (Supplementary Table S3). After annotation of these 49 transcripts on the human genome, we found that only 12 of these were true lncRNAs. The remaining 37 overlapped with protein coding genes, usually in the UTR regions. We used PCR analysis to examine 3 lncRNA transcripts that are located within annotated UTR regions of protein-coding genes in the human genome, but not in the bovine genome. Based on the presence or absence of PCR product after amplification with one primer located in the protein-coding gene and the other located in the hypothetical lncRNA from bovine cDNA, only mstrg.14819 could be considered as a lncRNA transcript distinct from the nearby protein-coding transcript. Among 12 conserved differentially expressed lncRNAs, we selected 5 (highlighted in yellow in Table 2), based on their expression levels, to investigate the intron-exon structure of the full transcripts. We chose to focus on these candidates as they were potentially important for development and differentiation of RPE. All were confirmed by PCR to be expressed in native bovine RPE.

### 3.3. Experimental Verification of Novel Bovine and Human lncRNAs.

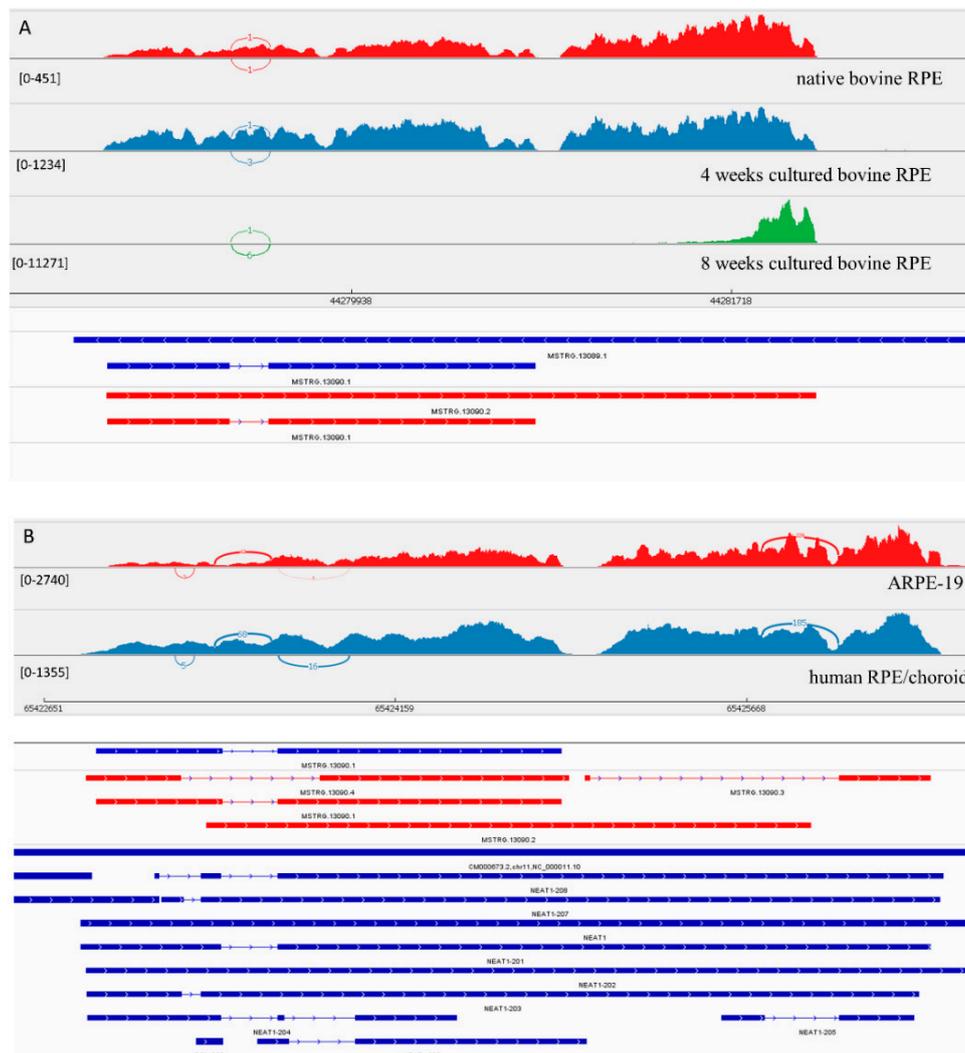
For experimental confirmation, essential for lncRNA studies, we selected 13 bovine transcripts with expression levels ranging from 0.6 to 400 Transcripts Per Kilobase Million (TPM) in native bovine RPE. All were confirmed as being expressed using quantitative PCR (Q-PCR), with Pearson correlation coefficient  $R = 0.9$  (Supplementary Table S4.). Conserved lncRNAs were in syntenic regions of the genomes, as expected by the parameters of the applied method. One of such regions contains newly annotated NEAT1 and MALAT1 genes in bovine genome (Supplementary Figure S2).

**Table 2.** lncRNAs conserved between human and bovine genomes and expressed in native RPE.

Gene Name	Native Retina/RPE			Bovine 4 Weeks Culture/Fresh RPE			Bovine 8 Weeks Culture/Fresh RPE			ARPE19 Cells 4D/4M			Human
	Base Mean	Fold Change	padj	Base Mean	Fold Change	padj	Base Mean	Fold Change	padj	Base Mean	Fold Change	padj	Annotation
MSTRG.13090	6175.61	2.43	$7.45 \times 10^{-31}$	9016.25	1.87	$8.01 \times 10^{-15}$	3028.53	1.42	$1.01 \times 10^{-3}$	0.64	0.12		NEAT1
MSTRG.13096	5.99	1.64	$2.18 \times 10^{-3}$	47.69	2.79	$2.44 \times 10^{-6}$	333.50	6.99	$2.04 \times 10^{-43}$	0.64	1.31		NEAT1
MSTRG.5066	2160.63	7.28	$1.38 \times 10^{-49}$	73.54	-3.38	$1.15 \times 10^{-11}$	30.64	-3.65	$5.59 \times 10^{-6}$	166.77	-4.29	$9.70 \times 10^{-42}$	linc00982
MSTRG.5821	5222.03	-0.52	$4.94 \times 10^{-4}$	351.11	-0.21	$5.20 \times 10^{-1}$	111.81	-1.34	$3.96 \times 10^{-4}$	2399.04	-0.22	$6.12 \times 10^{-3}$	TUG1
MSTRG.2517	212.94	-2.91	$5.43 \times 10^{-5}$	555.27	-1.79	$9.19 \times 10^{-10}$	239.46	-1.70	$5.02 \times 10^{-10}$	1087.27	-0.15	$2.24 \times 10^{-1}$	LINC000094
MSTRG.1397	55.94	0.21	$6.69 \times 10^{-1}$	5.07	-4.88	$1.55 \times 10^{-3}$	18.75	-5.98	$3.84 \times 10^{-6}$	15.11	-2.33	$5.31 \times 10^{-3}$	OTX2-AS1
MSTRG.1909	130.71	0.29	$2.35 \times 10^{-1}$	17.39	-1.98	$3.23 \times 10^{-2}$	28.53	-4.03	$4.62 \times 10^{-6}$	45.83	7.83	$4.05 \times 10^{-13}$	LINC01833
MSTRG.14528	599.24	-1.30	$7.77 \times 10^{-16}$	153.68	-1.22	$9.62 \times 10^{-5}$	88.45	-0.08	$8.86 \times 10^{-1}$	0.00			MAGI2-AS
MSTRG.5719	236.41	-3.93	$6.49 \times 10^{-20}$	357.87	-5.51	$5.83 \times 10^{-46}$	151.93	-5.61	$5.85 \times 10^{-20}$	0.00			RPS27P25-202
MSTRG.19295	56.36	0.08	$7.79 \times 10^{-1}$	674.42	1.22	$1.00 \times 10^{-3}$	20.14	2.93	$2.25 \times 10^{-5}$	0.50	0.48		FTX
MSTRG.9718	54.45	-0.61	$6.41 \times 10^{-1}$				76.35	5.31	$6.25 \times 10^{-7}$	2239.60	-12.44	$1.47 \times 10^{-42}$	MEG3
MSTRG.18250	23.79	-0.25	$8.50 \times 10^{-1}$				215.43	2.81	$5.71 \times 10^{-12}$	0.00			ZFAS1

### 3.3.1. LncRNA Gene: *NEAT1*

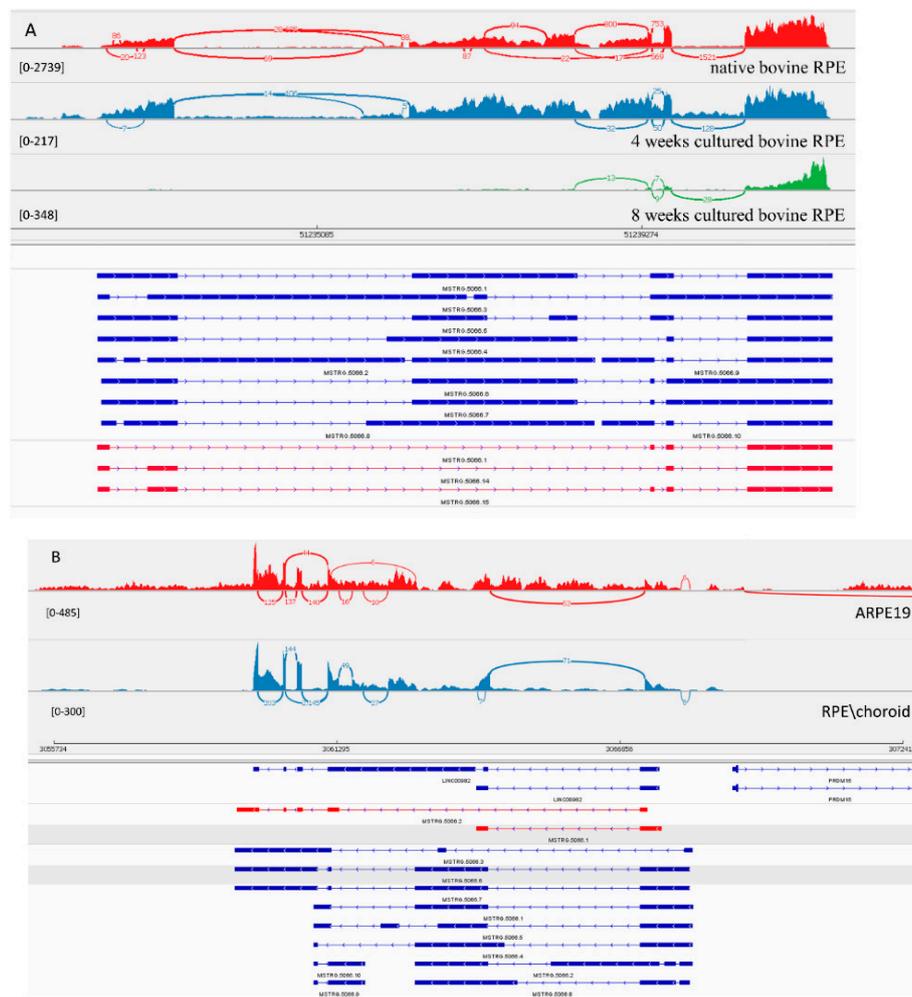
*MSTRG.13090.1*, the homolog of *NEAT1* (Nuclear Paraspeckle Assembly Transcript 1) in bovine RPE, has its expression level increased under culture conditions. *MSTRG.13089.1* is a longer homolog of *NEAT1* (Figure 2). The coverage of this transcript changed in cultured bovine RPE cells after 8 weeks in culture. There was an increased coverage of the 3' region of the *MSTRG.13089.1* transcript, suggesting the presence of a shorter novel isoform. Also, similar to the human transcripts, we suggest that in native bovine RPE cell these isoforms also were present. Experimentally in ARPE-19 cells, we confirmed the presence of 3 isoforms, that have not been previously annotated in Ensembl.



**Figure 2.** View of *NEAT1* from the Integrated Genome Viewer (IGV). IGV-Sashimi plots showing the read coverage and transcript isoforms. The histograms represent the sum of the aligned sequencing reads along the genome. Numbers in square brackets indicate coverage range. Junction reads are plotted as arcs, and the number of reads aligned to the junction spanning the exons connected by the arc is indicated. (A) Sashimi plots from bovine genome; red histogram represents the coverage for native bovine RPE, blue is 4-week cultured bovine RPE, and green is 8 weeks cultured bovine RPE. Below the histograms are shown the predicted isoforms (blue), experimentally verified in this study (red). (B) Sashimi plots from human genome; red histogram represents coverage for ARPE-19 cells and blue represents the native human RPE/choroid. Below the histograms are shown the predicted isoforms (blue), experimentally verified in this study (red), and Ensembl genome browser isoforms (blue).

### 3.3.2. LncRNA Gene: *MSTRG.5066*

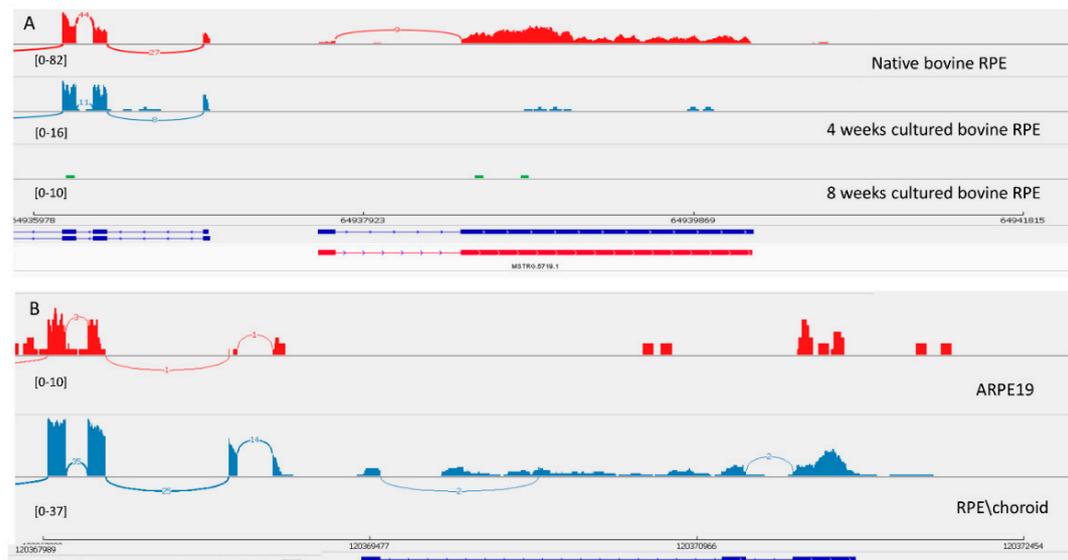
*MSTRG.5066* is homolog of human *linc00982* gene. Expression of transcripts from this gene were dramatically decreased upon explantation and in primary culture over time. This fact, and the conservation of this lincRNA, suggests a potential functional role of *linc00982* in the differentiation process of RPE. For this transcript, 10 isoforms were predicted computationally. In native bovine RPE we could clone only 3 shorter isoforms with 4 and 5 exons (Figure 3A). With such a complicated splicing pattern, Stringtie over-annotated its transcripts. In the human genome, 2 isoforms were confirmed with 5 and 2 exons (Figure 3B). Both matched 100% to previously annotated transcripts in Ensembl. Inexplicably, we could not amplify the transcript from bovine native RPE containing an exon predicted to belong in the middle of the transcript, and there was no PCR product from either the 5' or 3' end of the transcript containing this long exon.



**Figure 3.** View of *MSTRG.5066* from the Integrated Genome Viewer (IGV). IGV-Sashimi plots showing the read coverage and transcript isoforms. The histograms represent the sum of the aligned sequencing reads along the genome. Numbers in square brackets indicate coverage range. Junction reads are plotted as arcs, and the number of reads aligned to the junction spanning the exons connected by the arc is indicated. (A) Sashimi plots from bovine genome; red histogram represents the coverage for native bovine RPE, blue is 4-week cultured bovine RPE, and green is 8 weeks cultured bovine RPE. Below the histograms are shown the predicted isoforms (blue), experimentally verified in this study (red). (B) Sashimi plots from human genome; red histogram represents coverage for ARPE-19 cells and blue represents the native human RPE/choroid. Below the histograms are shown the GenBank or predicted isoforms (blue) and experimentally verified isoforms in this study (red).

### 3.3.3. LncRNA Gene: *MSTRG.5719*

Transcript *MSTRG.5719* was originally found in the bovine genome and was experimentally confirmed as a 100% match to the Stringtie prediction. This transcript also decreased in level of expression in culture. Interestingly, *MSTRG.5719* is located next to the splicing factor *MSI1* that controls photoreceptor-specific exons in retina [43]. It seems that the exon-intron structure of this transcript is different in the human genome (Figure 4). This transcript is not expressed in ARPE-19 cells, but in the human genome there is an annotated *RPS27P25-202* transcript that is homologous to the bovine transcript.



**Figure 4.** View of *MSTRG.5719* from the Integrated Genome Viewer (IGV). IGV-Sashimi plots showing the read coverage and transcript isoforms. The histograms represent the sum of the aligned sequencing reads along the genome. Numbers in square brackets indicate coverage range. Junction reads are plotted as arcs, and the number of reads aligned to the junction spanning the exons connected by the arc is indicated. (A) Sashimi plots from bovine genome; red histogram represents the coverage for native bovine RPE, blue is 4-week cultured bovine RPE, and green is 8 weeks cultured bovine RPE. Below the histograms are shown the predicted isoforms (blue), experimentally verified in this study (red). (B) Sashimi plots from human genome; red histogram represents coverage for ARPE-19 cells and blue represents the native human RPE/choroid. Below the histograms are shown the Ensembl genome browser isoform (blue).

### 3.3.4. LncRNA Gene: *MSTRG.1909* (*Linc1833*)

The *MSTRG.1909* transcript located in close proximity to the *SIX3* gene plays an important role in eye development. Expression of this transcript, as well as of *SIX3*, decreased in primary RPE culture compared with fresh native bovine RPE. We were able to confirm only the short isoform in human. This is likely to be an example of a lncRNA gene with a fast rate of evolution of exon-intron structure.

### 3.3.5. LncRNA Gene: *MSTRG.2517* (*Linc00094*)

*MSTRG.2517* had a similar pattern of diminishing expression in primary culture as *linc1833*, *MSTRG.5719*, and *linc00982*. *Linc00094/BRD3OS-202* is a convergent ncRNA that lies in the promoter region of the *BRD3* gene in the human genome. The *linc00094* isoform *BRD3OS-202* was confirmed as being expressed in ARPE-19 cells by cloning of the full transcript, but the *BRD3OS-205* transcript, coding a small protein of 84 amino acids, was not. We were unable to amplify this transcript from bovine RPE possibly because of its low expression.

#### 4. Discussion

In this study we found 12 multi-exon conserved RPE expressed lncRNAs between the human and bovine genomes. Many have known functions, but further analysis is needed for functional characterization of RPE-specific lncRNA isoforms. All the discussed lncRNAs significantly change their expression level under culture conditions, indicating their possible role in differentiation or the maintenance of the unique function of RPE cells. The strategy we have utilized could be used in future studies to annotate functional lncRNAs in other mammalian genomes.

The low number of conserved lncRNAs we found is in good agreement with current research [44,45], however a greater level of conservation is expected between primates where there are important evolutionary advances such as the macular region of the primate retina. Only 12 out of 700 studied lncRNAs are evolutionary conserved and differentially expressed in RPE, this suggests that less than 5% of lncRNA genes are conserved in RPE in the human–bovine comparison. This estimate is close to the estimate of conserved lncRNA genes that can be traced back to the origin of tetrapods (>300 million years ago) [23]. Another study produced a large fraction of conserved lncRNAs: for 1898 human lincRNAs expressed in nine tissues across multiple individuals, orthologous lincRNA transcripts were found for 39% in the cow, 38% in the mouse, and 35% in the rat [25]. It should be noted that estimates of fractions of conserved lncRNA genes depend on different conditions of experiments and even different definitions of conserved lncRNA genes. In any case, the comparison of various estimates are suggestive of the evolution of primate-specific lncRNAs to accommodate the needs of the new eye design.

Our study has some limitations. For example, lncRNAs with a single exon were not considered for analysis as they are considered unreliable, following previous studies [2,23,27,46], and conservation was assessed based on whole genome alignment with stringent identity present. This criterion could exclude some lncRNAs that were less conserved at the nucleotide level. In this regard, it has been suggested that the secondary and tertiary structure of lncRNAs could be more important for lncRNA function rather than merely the primary sequence of nucleotides alone [47,48].

Identified conserved lncRNAs are expressed in many tissues other than the RPE, but some isoforms might have tissue-specific expression, such as the shorter NEAT isoforms in human RPE. RPE-specific isoforms indicate distinct roles in the eye. Differences in the intron–exon structure of conserved lncRNAs suggest that there might be faster evolution of these regions, when compared to protein-coding genes.

RPE cells are important in the absorption of excess light, transport of nutrients and ions, the retinoid visual cycle, and in the phagocytosis of photoreceptor outer segments. These broad functions of RPE in the eye play a major role in determining its unique gene expression. We found a small set of conserved lncRNAs that could be involved in signature RPE functions conserved across mammals. Our findings demonstrate that the vast majority of lncRNA genes expressed in RPE are not evolutionarily conserved in mammalian genomes, nevertheless there may be a few conserved lncRNAs across various mammalian taxa that could be crucially important for eye function and development. However, the fraction of conserved lncRNAs in the overall pool of lncRNA found in RPE appeared to be very small, perhaps reflecting fast and flexible adaptation of the mammalian eye to various environmental conditions.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4425/10/3/205/s1>: Figure S1: IGV–Sashimi plot; Figure S2: IGV, syntenic region; Table S1: RPE lncRNAs classification; Table S2: RPE lncRNAs Rfam; Table S3: Differentially expressed RPE lncRNAs; Table S4: qPCR confirmation of bovine lncRNA expression; Supplementary file 1.GFT bovine; Supplementary file 2.GFT human.

**Author Contributions:** Author Contributions: Conceptualization, O.A.P., I.B.R., E.P. and T.M.R.; Methodology, O.A.P., E.P., I.B.R., and G.N.; Software, I.B.R., V.N.B., G.N.; Validation, O.A.P., I.B.R., G.N. and V.N.B.; Formal Analysis, O.A.P., I.B.R., and V.N.B.; Investigation, O.A.P., I.B.R., E.P. and T.M.R.; Resources, O.A.P., W.S., I.B.R., V.N.B. and G.N.; Data Curation, O.A.P., W.S., I.B.R., V.N.B. and G.N.; Writing—Original Draft Preparation, O.A.P., I.B.R., G.N., V.N.B., E.P. and T.M.R.; Writing—Review and Editing, O.A.P., I.B.R., W.S., V.N.B., G.N., E.P. and T.M.R.

**Funding:** This research was supported by the Intramural Research Program of the National Eye Institute, NIH, NIH and National Library of Medicine, National Institutes of Health.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D.G.; et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **2012**, *22*, 1775–1789. [[CrossRef](#)] [[PubMed](#)]
2. Cabili, M.N.; Trapnell, C.; Goff, L.; Koziol, M.; Tazon-Vega, B.; Regev, A.; Rinn, J.L. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **2011**, *25*, 1915–1927. [[CrossRef](#)] [[PubMed](#)]
3. Bertone, P.; Stolc, V.; Royce, T.E.; Rozowsky, J.S.; Urban, A.E.; Zhu, X.; Rinn, J.L.; Tongprasit, W.; Samanta, M.; Weissman, S.; et al. Global identification of human transcribed sequences with genome tiling arrays. *Science* **2004**, *306*, 2242–2246. [[CrossRef](#)] [[PubMed](#)]
4. Mattick, J.S.; Makunin, I.V. Non-coding RNA. *Hum. Mol. Genet.* **2006**, *15*, R17–R29. [[CrossRef](#)]
5. Mercer, T.R.; Dinger, M.E.; Mattick, J.S. Long non-coding RNAs: Insights into functions. *Nat. Rev. Genet.* **2009**, *10*, 155–159. [[CrossRef](#)] [[PubMed](#)]
6. Ponjavic, J.; Ponting, C.P.; Lunter, G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* **2007**, *17*, 556–565. [[CrossRef](#)] [[PubMed](#)]
7. Ponting, C.P.; Belgard, T.G. Transcribed dark matter: Meaning or myth? *Hum. Mol. Genet.* **2010**, *19*, R162–R168. [[CrossRef](#)] [[PubMed](#)]
8. Ponting, C.P.; Oliver, P.L.; Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **2009**, *136*, 629–641. [[CrossRef](#)]
9. Khalil, A.M.; Guttman, M.; Huarte, M.; Garber, M.; Raj, A.; Rivea Morales, D.; Thomas, K.; Presser, A.; Bernstein, B.E.; van Oudenaarden, A.; et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11667–11672. [[CrossRef](#)]
10. Robinson, R. Dark matter transcripts: Sound and fury, signifying nothing? *PLoS Biol.* **2010**, *8*, e1000370. [[CrossRef](#)]
11. Van Bakel, H.; Hughes, T.R. Establishing legitimacy and function in the new transcriptome. *Brief Funct. Genom. Proteom.* **2009**, *8*, 424–436. [[CrossRef](#)] [[PubMed](#)]
12. Marques, A.C.; Ponting, C.P. Catalogues of mammalian long noncoding RNAs: Modest conservation and incompleteness. *Genome Biol.* **2009**, *10*, R124. [[CrossRef](#)] [[PubMed](#)]
13. Siepel, A.; Bejerano, G.; Pedersen, J.S.; Hinrichs, A.S.; Hou, M.; Rosenbloom, K.; Clawson, H.; Spieth, J.; Hillier, L.W.; Richards, S.; et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **2005**, *15*, 1034–1050. [[CrossRef](#)] [[PubMed](#)]
14. Guttman, M.; Amit, I.; Garber, M.; French, C.; Lin, M.F.; Feldser, D.; Huarte, M.; Zuk, O.; Carey, B.W.; Cassady, J.P.; et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **2009**, *458*, 223–227. [[CrossRef](#)]
15. Managadze, D.; Rogozin, I.B.; Chernikova, D.; Shabalina, S.A.; Koonin, E.V. Negative correlation between expression level and evolutionary rate of long intergenic noncoding RNAs. *Genome Biol. Evol.* **2011**, *3*, 1390–1404. [[CrossRef](#)] [[PubMed](#)]
16. Huttenhofer, A.; Vogel, J. Experimental approaches to identify non-coding RNAs. *Nucl. Acids Res.* **2006**, *34*, 635–646. [[CrossRef](#)] [[PubMed](#)]
17. Amaral, P.P.; Dinger, M.E.; Mattick, J.S. Non-coding RNAs in homeostasis, disease and stress responses: An evolutionary perspective. *Brief Funct. Genom.* **2013**, *12*, 254–278. [[CrossRef](#)]
18. Carninci, P.; Kasukawa, T.; Katayama, S.; Gough, J.; Frith, M.C.; Maeda, N.; Oyama, R.; Ravasi, T.; Lenhard, B.; Wells, C.; et al. The transcriptional landscape of the mammalian genome. *Science* **2005**, *309*, 1559–1563.
19. Liu, J.; Gough, J.; Rost, B. Distinguishing protein-coding from non-coding RNAs through support vector machines. *PLoS Genet* **2006**, *2*, e29. [[CrossRef](#)]

20. Ji, Z.; Song, R.; Regev, A.; Struhl, K. Many lincRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *Elife* **2015**, *4*, e08890. [[CrossRef](#)]
21. Managadze, D.; Lobkovsky, A.E.; Wolf, Y.I.; Shabalina, S.A.; Rogozin, I.B.; Koonin, E.V. The vast, conserved mammalian lincRNome. *PLoS Comput. Biol.* **2013**, *9*, e1002917. [[CrossRef](#)] [[PubMed](#)]
22. Chodroff, R.A.; Goodstadt, L.; Sirey, T.M.; Oliver, P.L.; Davies, K.E.; Green, E.D.; Molnar, Z.; Ponting, C.P. Long noncoding RNA genes: Conservation of sequence and brain expression among diverse amniotes. *Genome Biol.* **2010**, *11*, R72. [[CrossRef](#)]
23. Necsulea, A.; Soumillon, M.; Warnefors, M.; Liechti, A.; Daish, T.; Zeller, U.; Baker, J.C.; Grutzner, F.; Kaessmann, H. The evolution of lincRNA repertoires and expression patterns in tetrapods. *Nature* **2014**, *505*, 635–640. [[CrossRef](#)]
24. Ulitsky, I.; Shkumatava, A.; Jan, C.H.; Sive, H.; Bartel, D.P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **2011**, *147*, 1537–1550. [[CrossRef](#)] [[PubMed](#)]
25. Washietl, S.; Kellis, M.; Garber, M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res.* **2014**, *24*, 616–628. [[CrossRef](#)] [[PubMed](#)]
26. Chernikova, D.; Managadze, D.; Glazko, G.V.; Makalowski, W.; Rogozin, I.B. Conservation of the Exon-Intron Structure of Long Intergenic Non-Coding RNA Genes in Eutherian Mammals. *Life (Basel)* **2016**, *6*, 27. [[CrossRef](#)] [[PubMed](#)]
27. Kapusta, A.; Feschotte, C. Volatile evolution of long noncoding RNA repertoires: Mechanisms and biological implications. *Trends Genet.* **2014**, *30*, 439–452. [[CrossRef](#)] [[PubMed](#)]
28. Mustafi, D.; Kevany, B.M.; Bai, X.; Maeda, T.; Sears, J.E.; Khalil, A.M.; Palczewski, K. Evolutionarily conserved long intergenic non-coding RNAs in the eye. *Hum. Mol. Genet.* **2013**, *22*, 2992–3002. [[CrossRef](#)]
29. Au, E.D.; Fernandez-Godino, R.; Kaczynski, T.J.; Sousa, M.E.; Farkas, M.H. Characterization of lincRNA expression in the human retinal pigment epithelium and differentiated induced pluripotent stem cells. *PLoS ONE* **2017**, *12*, e0183939. [[CrossRef](#)]
30. Wan, P.; Su, W.; Zhuo, Y. Precise long non-coding RNA modulation in visual maintenance and impairment. *J. Med. Genet.* **2017**, *54*, 450–459. [[CrossRef](#)]
31. Rapicavoli, N.A.; Poth, E.M.; Blackshaw, S. The long noncoding RNA RNCR2 directs mouse retinal cell specification. *BMC Dev. Biol.* **2010**, *10*, 49. [[CrossRef](#)] [[PubMed](#)]
32. Krol, J.; Krol, I.; Alvarez, C.P.; Fiscella, M.; Hierlemann, A.; Roska, B.; Filipowicz, W. A network comprising short and long noncoding RNAs and RNA helicase controls mouse retina architecture. *Nat. Commun.* **2015**, *6*, 7305. [[CrossRef](#)] [[PubMed](#)]
33. Chen, W.; Yang, S.; Zhou, Z.; Zhao, X.; Zhong, J.; Reinach, P.S.; Yan, D. The Long Noncoding RNA Landscape of the Mouse Eye. *Invest. Ophthalmol. Vis. Sci.* **2017**, *58*, 6308–6317. [[CrossRef](#)] [[PubMed](#)]
34. Samuel, W.; Jaworski, C.; Postnikova, O.A.; Kuttly, R.K.; Duncan, T.; Tan, L.X.; Poliakov, E.; Lakkaraju, A.; Redmond, T.M. Appropriately differentiated ARPE-19 cells regain phenotype and gene expression profiles similar to those of native RPE cells. *Mol. Vis.* **2017**, *23*, 60–89. [[PubMed](#)]
35. Whitmore, S.S.; Wagner, A.H.; DeLuca, A.P.; Drack, A.V.; Stone, E.M.; Tucker, B.A.; Zeng, S.; Braun, T.A.; Mullins, R.F.; Scheetz, T.E. Transcriptomic analysis across nasal, temporal, and macular regions of human neural retina and RPE/choroid by RNA-Seq. *Exp. Eye Res.* **2014**, *129*, 93–106. [[CrossRef](#)] [[PubMed](#)]
36. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)] [[PubMed](#)]
37. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [[CrossRef](#)]
38. Wucher, V.; Legeai, F.; Hedan, B.; Rizk, G.; Lagoutte, L.; Leeb, T.; Jagannathan, V.; Cadieu, E.; David, A.; Lohi, H.; et al. FEELnc: A tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucl. Acids Res.* **2017**, *45*, e57. [[CrossRef](#)]
39. Goecks, J.; Nekrutenko, A.; Taylor, J.; Galaxy, T. Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **2010**, *11*, R86. [[CrossRef](#)]
40. Prakash, A.; Tompa, M. Measuring the accuracy of genome-size multiple alignments. *Genome Biol.* **2007**, *8*, R124. [[CrossRef](#)]

41. Rogozin, I.B.; D'Angelo, D.; Milanesi, L. Protein-coding regions prediction combining similarity searches and conservative evolutionary properties of protein-coding sequences. *Gene* **1999**, *226*, 129–137. [[CrossRef](#)]
42. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)] [[PubMed](#)]
43. Murphy, D.; Ciepły, B.; Carstens, R.; Ramamurthy, V.; Stoilov, P. The Musashi 1 controls the splicing of photoreceptor-specific exons in the vertebrate retina. *PLoS Genet.* **2016**, *12*, e1006256. [[CrossRef](#)] [[PubMed](#)]
44. Hutchinson, J.N.; Ensminger, A.W.; Clemson, C.M.; Lynch, C.R.; Lawrence, J.B.; Chess, A. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genom.* **2007**, *8*, 39. [[CrossRef](#)] [[PubMed](#)]
45. Johnsson, P.; Lipovich, L.; Grander, D.; Morris, K.V. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim. Biophys. Acta* **2014**, *1840*, 1063–1071. [[CrossRef](#)] [[PubMed](#)]
46. Clark, M.B.; Johnston, R.L.; Inostroza-Ponta, M.; Fox, A.H.; Fortini, E.; Moscato, P.; Dinger, M.E.; Mattick, J.S. Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **2012**, *22*, 885–898. [[CrossRef](#)] [[PubMed](#)]
47. Guttman, M.; Rinn, J.L. Modular regulatory principles of large non-coding RNAs. *Nature* **2012**, *482*, 339–346. [[CrossRef](#)]
48. Novikova, I.V.; Hennelly, S.P.; Sanbonmatsu, K.Y. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucl. Acids Res.* **2012**, *40*, 5034–5051. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).