2022-03

Single molecule, long-read Apoer2 sequencing identifies conserved and species-specific splic...

Genomics 114 (2022) 110318

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Single molecule, long-read *Apoer2* sequencing identifies conserved and species-specific splicing patterns

Christina M. Gallo [a,b], Adam T. Labadorf [c,d], Angela Ho [a,b,*], Uwe Beffert [b]

[a] Department of Pharmacology & Experimental Therapeutics, Boston University School of Medicine, United States of America
[b] Department of Biology, Boston University, United States of America
[c] Bioinformatics Program, Boston University, United States of America
[d] Department of Neurology, Boston University School of Medicine, United States of America

ARTICLE INFO

Keywords:
APOER2
LRP8
Alternative splicing

ABSTRACT

Apolipoprotein E receptor 2 (Apoer2) is a synaptic receptor in the brain that binds disease-relevant ligand Apolipoprotein E (Apoe) and is highly alternatively spliced. We examined alternative splicing (AS) of conserved *Apoer2* exons across vertebrate species and identified gain of exons in mammals encoding functional domains such as the cytoplasmic and furin inserts, and loss of an exon in primates encoding the eighth LDLa repeat, likely altering receptor surface levels and ligand-binding specificity. We utilized single molecule, long-read RNA sequencing to profile full-length *Apoer2* isoforms and identified 68 and 48 unique full-length *Apoer2* transcripts in the mouse and human cerebral cortex, respectively. Furthermore, we identified two exons encoding protein functional domains, the third EGF-precursor like repeat and glycosylation domain, that are tandemly skipped specifically in mouse. Our study provides new insight into *Apoer2* isoform complexity in the vertebrate brain and highlights species-specific differences in splicing decisions that support functional diversity.

1. Introduction

The human genome contains about 20,000 protein coding genes [1], comparable to much simpler organisms such as worms or flies. However, over 100,000 unique proteins are thought to be synthesized in humans [2]. One major determinant in proteome diversity is RNA alternative splicing (AS). In AS, different combinations of splice sites are used during pre-mRNA splicing leading to multiple unique RNA isoforms from an individual gene. AS can occur in different forms, including use of alternative 5′ or 3′ splice-sites, cassette-exon inclusion or skipping and intron retention [3]. 92–94% of human genes undergo AS, with 86% of those genes demonstrating a minor isoform frequency of 15% or more [4]. This greatly expands the proteome and generates complexity from individual genes, as RNA isoforms can possess altered stability, localization, translation competency or coding sequence [5,6]). Notably, about 15% of disease causing mutations in humans lead to mRNA splicing defects [7,8].

AS is more prevalent in higher eukaryotes compared to lower eukaryotes, particularly in vertebrates [9,10]. The prevalence of cassette exon skipping increases further up the evolutionary tree, suggesting that exon skipping may be the type of AS that contributes most to phenotypic complexity [10,11]. In addition, splicing regulatory protein families have expanded further up the evolutionary tree, including families such as the serine/arginine proteins (SR proteins) and the heterogenous nuclear ribonucleoproteins (hnRNPs) family [12].

In exons themselves, a correlation has been observed between the borders of exons and individual domains at the protein level [13] that becomes stronger as organismal complexity increases [14]. This modular design, likely produced through exon shuffling [15], allows for diversification at the protein level with the addition or subtraction of unique protein domains, provided AS events do not disrupt the protein open reading frame. Interestingly, many modular proteins generated through exon shuffling are involved in mediating cell-cell or cell-extracellular matrix interactions, including members of the Low Density Lipoprotein Receptor (LDLR) family [16,17]. Conserved alternative exons between mice and humans are also enriched for genes expressed in the brain [18].

Apolipoprotein E receptor 2 (Apoer2), official gene name *Lrp8*, is a modular type I transmembrane receptor of the LDLR family that is enriched in the vertebrate brain [19]. Mouse Apoer2 regulates the layering of the neocortex in development [20] and long term potentiation (LTP) in the adult brain [21]. Importantly, Apoer2 is an

* Corresponding author at: 24 Cummington Mall, Boston, MA 02215, United States of America.
E-mail addresses: aho1@bu.edu, ub1@bu.edu (A. Ho).

https://doi.org/10.1016/j.ygeno.2022.110318
Received 30 September 2021; Received in revised form 11 February 2022; Accepted 15 February 2022
Available online 19 February 2022

Apolipoprotein E (Apoe) receptor [19]. Apoe is a secreted glycoprotein that acts as a cholesterol carrier and signaling molecule [22–24] and serves as an isoform-specific risk factor for sporadic Alzheimer's disease [25,26]. Interestingly, human and mouse *Apoer2* are highly alternatively spliced in the brain [27,28], potentially altering the receptor's ability to interact with critical ligands like Apoe. Some *Apoer2* AS events demonstrate conservation across multiple vertebrate species, including chickens and mice [19,29], and exhibit isoform-specific functions, including ligand binding properties [30–32].

Human APOER2 consists of five functional domains [19] that contain distinct modular protein domains. The first is an amino terminal ligand binding domain made up of two types of binding repeats: LDL receptor type A (LDLa) repeats and epidermal growth factor (EGF) precursor-like repeats. The second functional domain is the YWTD β-propeller domain that facilitates ligand release and receptor recycling after endocytosis [33]. The third domain is an O-linked sugar domain critical for receptor glycosylation and surface expression [34]. The hydrophobic transmembrane region makes up the fourth functional domain. Lastly, the fifth domain is a short cytoplasmic tail containing an NPXY motif critical for clathrin-mediated endocytosis and binding of adaptor proteins [35–37]. The cytoplasmic tail of Apoer2 also includes a proline rich 59 amino acid insert that is essential for Reelin-induced enhancement of LTP in mice [30] and is unique to Apoer2 [19]. These individual APOER2 protein domains are largely encoded by distinct cassette exons that are in phase, allowing for cassette exon skipping events that preserve the open reading frame and generate multiple unique proteins [38]. Furthermore, conservation of several of these AS events across species, such as chicken, mice and humans [27,29,39,40], suggests these events may have important biological function since the splicing event has avoided negative selection [10].

Apoer2 was originally thought to be two distinct receptors of the LDLR family due to the presence of an eighth LDLa repeat that is present in chickens and mice but not in humans [19,29,41]. Further analysis revealed that this exon was lost in monkeys and humans due to mutation of the 3′ splice site [38]. This indicates that there are evolutionary changes in *Apoer2* across vertebrates, suggesting *Apoer2* has undergone evolutionary selection in the vertebrate brain. Other exons encoding functional protein domains in Apoer2 have been identified as alternatively spliced, including the cytoplasmic insert in mice and humans [29,30,40,42], the 39-nucleotide exon that introduces a furin cleavage site in mice and humans [29,43] and the glycosylation domain in mice and humans [27,34,44], among others. However, a systematic tracing of *Apoer2* splicing events across the vertebrate lineage is lacking. It is unclear how many distinct *Apoer2* isoforms are physiologically produced and how often individual *Apoer2* cassette exon splicing events occur in tandem across the entire length of the *Apoer2* transcript.

In this study, we have performed a direct comparison of *Apoer2* alternatively spliced exons across vertebrates, including zebrafish, chickens, rabbits, mice, monkeys and humans. We demonstrate that AS of *Apoer2* increases with evolutionary complexity and show the loss of the eighth LDLa repeat from mice to monkeys. Our analyses indicate the exon encoding a furin cleavage site appears to have evolved between chicken and rabbits, which coincides with the appearance of the alternatively spliced cytoplasmic insert. Furthermore, to investigate coordinated splicing of *Apoer2* cassette exons, we have utilized single molecule, long-read sequencing to profile splicing events across *Apoer2* in the cerebral cortex of mice and humans. We demonstrate a rich complexity of AS combinations across both humans and mice, some of which are conserved. Conservation of AS events suggests they have significant effects at the functional level, which has already been shown for AS of the cytoplasmic insert in mouse Apoer2 [30]. Therefore, it is likely that other AS events or AS combinations in Apoer2 also endow distinct functional properties at the protein level in the brain, such as modulating Apoe binding. We also observed and validated a tandem splicing event of the exons encoding the third EGF precursor-like repeat and the glycosylation domain in mice, but not humans, suggesting differential splicing regulation exists across the two species.

2. Materials and methods

2.1. Phylogenetic and isoform analysis

All annotated *Lrp8* RNA isoforms were downloaded from each NCBI and Ensembl gene databases for *Danio rerio* (zebrafish), *Gallus gallus* (chicken), *Oryctolagus cuniculus* (rabbit), *Mus musculus* (mouse), *Macaca fascicularis* (cynomolgus monkey/macaque) and *Homo sapiens* (human). Sequences were analyzed and compared to their corresponding reference genome in NCBI and exons annotated manually using SnapGene software. To examine homology between species at the protein level, the longest coding isoform was selected for each species and the corresponding protein sequence was aligned with that of all other species. NCBI's Constraint-based Multiple Alignment Tool (COBALT) was used to generate protein alignment and phylogenetic tree using the default COBALT parameters [45]. To visualize protein alignment, FASTA output was input into ExPASY's BoxShade tool with an output of RTF_new (Fig. S1).

2.2. RNA isolation, cDNA synthesis and RT-PCR

Total RNA was isolated from the brains of four adult wild-type zebrafish brains using TRIzol reagent (Invitrogen) according to manufacturer instructions. Genomic DNA depletion was performed immediately after using DNase I in Qiagen's RNeasy Kit according to manufacturer instructions. Total RNA from the whole brain of chicken (CR-201), rabbit (TR-201) and cynomolgus monkey (KR-201) was purchased from Amsbio. Total RNA from the cerebral cortex of mouse (636661) and human (636561) was purchased from Takara Biosciences. RNA was quantified using a NanoDrop spectrophotometer. cDNA synthesis was carried out using Superscript IV First-Strand Synthesis System (Invitrogen). 1.5 μg of RNA was used as input for annealing of oligo-dT primers. Final clean-up was performed with 1 μL of RNase H.

RT-PCR was performed using either PrimeStar Polymerase (Takara Biosciences) or Q5 Hot-Start High-Fidelity Polymerase (NEB) depending on optimization. Optimal annealing temperature was determined using a gradient thermocycler. Primers were designed and screened for specificity using NCBI's Primer-BLAST [46] (Supplemental Dataset S2).

RT-PCR reactions were separated out by agarose gel electrophoresis in 1× TAE Buffer. Ethidium bromide was used for visualization. All DNA bands generated from RT-PCR reactions were individually excised and purified using BioBasic's Plasmid DNA Isolation Miniprep Kit according to manufacturer protocol for agarose gel DNA purification. Each product was then individually cloned for sequencing using Invitrogen's Zero Blunt TOPO PCR Cloning Kit. Two to ten clones were picked per individual RT-PCR band and verified by sanger sequencing with vector specific primer M13F. Sequencing analysis was carried out using SnapGene and NCBI BLAST [47].

2.3. Pacific Biosciences library preparation and long read sequencing

Similar to the methodology used in Treutlein, et al. [48], gene specific first strand cDNA synthesis was first performed targeting *Lrp8*, followed by further *Lrp8* specific RT-PCR. Briefly, 1 μg of RNA per sample was incubated with 1 μL of 10 mM dNTPs and 1 μL *Lrp8* reverse primer (Human REV: 5′- TCA GGG TAG TCC ATC ATC TTC AAG GC-3′, Mouse REV: 5′-TCA GGG CAG TCC ATC ATC TTC AAG AC- 3′) in a 10 μL reaction for 5 min at 65 °C. The reaction was then cooled on ice for at least 2 min. The reaction was then supplemented to 20 μL with 5× First Strand Buffer (Invitrogen), DTT, Superase In RNAse Inhibitor and SuperScript RTase followed by first strand synthesis: 1 h at 55 °C, 15 min at 70 °C, hold at 4 °C. 2 units (1 μL) of RNase H was then added prior to a 20-min incubation at 37 °C.

RT-PCR reaction was performed in multiple 50 μL reactions and

pooled for subsequent AMPure XP SPRI clean-up. RT-PCR mix utilized: 5× Q5 Buffer (NEB), 5 μg BSA, 1 M Betaine, 3% DMSO, 0.2 mM dNTP, 0.2 μM FWD primer (Human FWD: 5′-ATG GGC CTC CCC GAG CC -3′, Mouse FWD: 5′- CTA TTA TGG GCC GCC CAG AAC TGG G- 3′), 0.2 μM REV primer (Human REV, Mouse REV), Q5 High Fidelity Polymerase, and 2 μL of first strand synthesis reaction per 50 μL. Cycle utilized for human RT-PCR is as follows: 98 °C 2 min; 30 cycles of: 98 °C- 10 s, 64 °C- 30 s, 65 °C- 1 min 40 s; 72 °C- 10 min, 4 °C hold. Cycle utilized for mouse RT-PCR is as follows: 98 °C 2 min; 30 cycles of: 98 °C- 10 s, 68.6 °C- 30 s, 72 °C- 2 min; 72 °C- 10 min, 4 °C hold.

RT-PCR reactions were then pooled and subject to 0.5× SPRI clean-up using AMPure XP beads, with a final elution using 20 μL nuclease-free water. To determine concentration and purity, cDNA amplicons were then quantified using both NanoDrop Spectrophotometer and Qubit Fluorometer. 1 μg of DNA amplicon was prepared for PacBio library preparation.

500 ng of each cDNA pool was prepared *via* the express template preparation kit 2.0 from Pacific Biosciences. Each sample underwent single-strand overhang removal, DNA damage repair, and end repair/a-tailing by standard methods. Barcoded overhang SMRTbell adapters were used for each sample in place of the standard unbarcoded SMRTbell adapters. After ligation, samples were pooled in an equimolar fashion to a total mass of 1000 ng per pool. Pooled libraries were cleaned and concentrated with 0.45× AMPure beads and underwent polymerase annealing and binding by standard methods. Primer v4 was used. Each library was loaded with Sequel loading kit 2.1 at 60pM. A 2-h pre-extension time and a 24-h movie was used. CCS and barcode demultiplexing were carried out by SMRTlink v8.0.

2.4. Processing of Pacific Biosciences single molecule, long-read sequencing data

SMRTlink v8.0 was used to process demultiplexed sequencing files. PacBio primers and *Lrp8* specific RT-PCR primers were clipped, primer concatemers were removed and sequences were clustered into polished high-quality isoforms using SMRTlink v8.0 pipeline with standard Iso-Seq parameters. High-quality isoforms were collapsed into unique isoforms using cDNA Cupcake (https://github.com/Magdoll/cDNA_Cupcake, v17.0.0) with a maximum 5′ and 3′ difference of 50 base pairs and without merging shorter isoforms. cDNA Cupcake was also used to generate isoform abundance using SMRTLink cluster reports and 5′ fragments were filtered out along with any isoforms with a full-length read count less than two. cDNA Cupcake generated fastq files were then reference corrected against the mouse (mm10) or human (hg38) genome using SQANTI3 v.2.0.0 [49].

2.5. Transcript annotation and homology analysis

Corrected transcript output from SQANTI3 [49] was parsed in R [50] to include only *Apoer2* transcripts. All transcripts were filtered based off the isoform FASTA file so that they contained the final sequence of *Apoer2* located in the last exon, just prior to the reverse primer, since the small length of this sequence prevented annotation as an individual exon. The intersection of isoforms after these two filtering steps was moved into further analysis. For exon annotation, the SQANTI3 gff3 output file was parsed for *Apoer2* exons, and unique exons were extracted and manually annotated using the Broad Institute's Integrative Genomics Viewer (IGV) [51]. Exons corresponding to individual isoforms were then assembled into a binary splice matrix in R, and exons were labeled using previous manual IGV annotation. *Apoer2* isoform sequences were deposited in NCBI's GenBank (see Supplementary Dataset S4 for accession numbers).

To determine homologous transcripts between mice and humans, homologous exons were matched based off previous Apoer2 protein alignment and exon annotation (Supplementary Dataset S3). Transcripts with exons outside of known homologous exons between the two species were filtered out. Remaining isoforms were matched between the two species in R based on having matching patterns of corresponding exons across the transcript.

2.6. Transcript mapping and exon level analysis

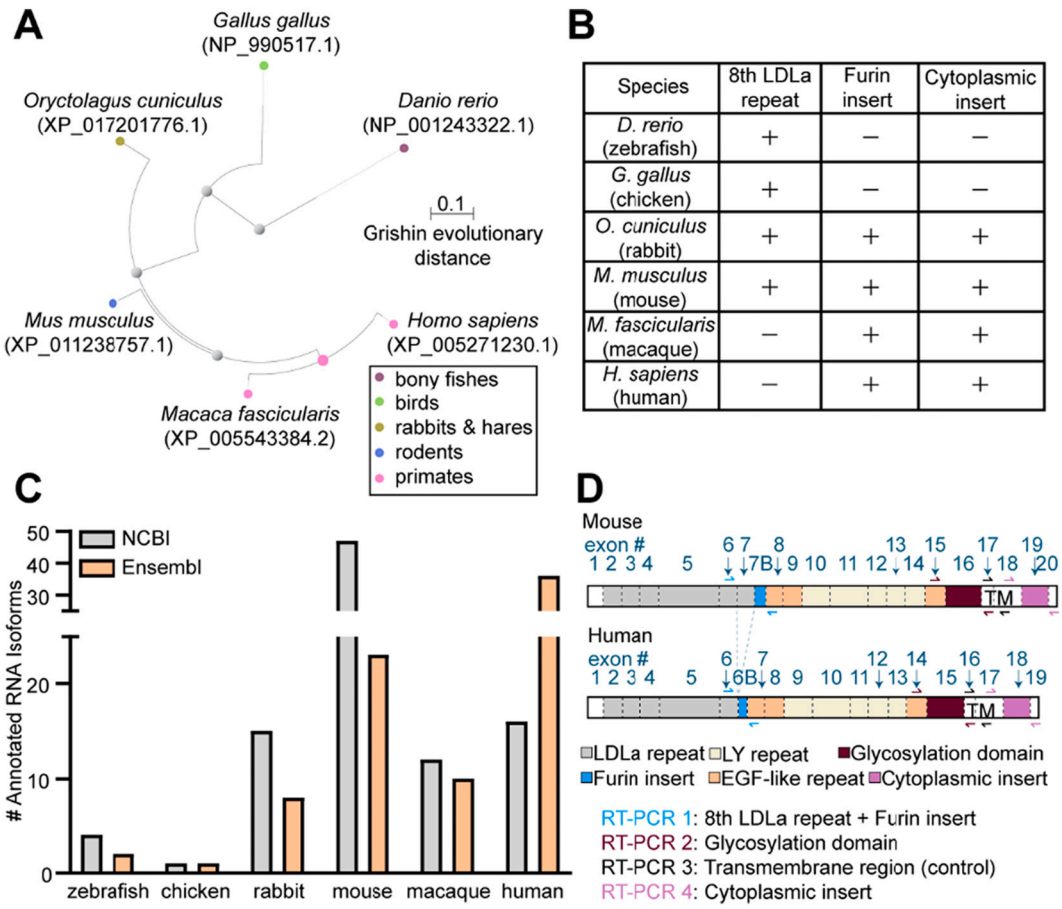
The binary splice matrix was assembled into a transcript matrix using ggplot2 [52] and publicly available code deposited in github from Flaherty, et al. [53]. Coincidence was calculated by counting the number of times an exon was spliced in at the same time as each other exon, weighted by transcript abundance and then divided by the total number of transcripts. Heatmaps and barplots were generated using ggplot2. Frequency spliced in was calculated by counting the number of times an exon was spliced in over all transcripts weighted by transcript abundance. Transcript length, cumulative frequency and parts of a whole annotation graphs were generated using GraphPad Prism v 9.2.0. Correlation matrices were generated using the binary splice matrix and R packages Hmsic [54] and corrplot [55] to analyze the Pearson correlation coefficient. A significance level of 0.01 was used.

3. Results

3.1. Apoer2 cassette exon skipping across vertebrates

To examine Apoer2 homology across vertebrates, we aligned full-length Apoer2 protein sequences from zebrafish, chicken, rabbit, mouse, macaque and human (Fig. S1). Fig. 1*A* demonstrates the phylogeny of Apoer2 as phenotypic complexity increases from zebrafish to humans. To understand how Apoer2 protein domains changed over evolution from zebrafish to humans, we mapped the known protein domains of human APOER2 onto the protein alignment to determine whether these domains were conserved in other species. We found that chickens and mice contain an eighth LDLa ligand binding repeat in their extracellular domain, while monkeys and humans do not (Fig. 1*B*), consistent with previous studies [29,38,40]. We also observed the presence of the eighth LDLa repeat in zebrafish and rabbits. *Apoer2* also contains a 39-nucleotide exon at the end of its LDLa repeats in mice and humans that encodes a cleavage site for the furin protease [29,43]. This 39-nucleotide exon appears to be present in both rabbits and macaques yet lacking in zebrafish and chickens. Lastly, Apoer2 contains a unique 59 amino acid proline rich insert in its cytoplasmic region that has been reported as present in mice but lacking in birds [29,56]. Based on our protein alignment, neither zebrafish nor chickens possess the cytoplasmic insert; however, it is apparent in vertebrates higher than the rabbit in this study.

To understand what is currently known about AS events in *Apoer2* across species, we curated all known *Apoer2* isoforms deposited in National Center for Biotechnology Information (NCBI) [57] and Ensembl [58] online databases. There is a general increase in the number of annotated *Apoer2* isoforms as evolutionary complexity increases (Fig. 1*C*). However, published work has demonstrated alternative isoforms of *Apoer2* do exist in chickens [29,41], which is not currently reflected in these annotation databases. This is perhaps due to the widespread usage of mice or human cells and tissue to study splicing and other biological processes resulting in greater sequencing coverage. To address this discrepancy, we analyzed the splicing patterns of known alternative exons in *Apoer2* including the exons encoding the eighth LDLa repeat, the furin cleavage site, the glycosylation domain and the cytoplasmic insert. We mapped the protein sequence of each species used for alignment back onto their corresponding RNA transcript to identify the homologous exons to target (Supplementary Dataset S1). RT-PCR primers were placed as indicated, flanking the alternative exons of interest, as demonstrated for mouse and human Apoer2 in Fig. 1*D*. RT-PCR reactions will yield one band for exon inclusion and an additional one for exon skipping, if it occurs. The eighth LDLa repeat and the furin cleavage site were captured using the same RT-PCR reaction, generating the



Species	8th LDLa repeat	Furin insert	Cytoplasmic insert
D. rerio (zebrafish)	+	–	–
G. gallus (chicken)	+	–	–
O. cuniculus (rabbit)	+	+	+
M. musculus (mouse)	+	+	+
M. fascicularis (macaque)	–	+	+
H. sapiens (human)	–	+	+

Fig. 1. Apoer2 protein and isoform diversity across vertebrates. (A) Phylogenetic tree of Apoer2 spanning from *Danio rerio* (zebrafish) to *Homo sapiens* (humans) displaying Grishin evolutionary distance. (B) Table indicating presence or absence of Apoer2 protein domains of interest across vertebrates based off protein reference annotations. (C) Bar graph depicting the number of annotated *Apoer2* isoforms for each vertebrate species in each the NCBI and Ensembl online databases. (D) Schematic of Apoer2 protein domains and corresponding coding exons in each mice and humans. RT-PCR primer schemes are indicated using arrows.

possibility of observing four bands indicative of inclusion and exclusion of each the exon encoding the eighth LDLa repeat and the exon encoding the furin cleavage site. Whole brain total RNA from zebrafish, chicken, rabbit and monkey and total RNA from the cerebral cortex of mice and humans were subject to oligo-dT based cDNA synthesis and RT-PCR targeting the *Apoer2* regions of interest. RT-PCR reactions were then visualized by gel electrophoresis (Fig. 2*A*–*B*) and all bands were confirmed by sequencing.

Zebrafish demonstrate constitutive inclusion of the eighth LDLa ligand binding repeat (Fig. 2*C*) and do not appear to have the furin insert as indicated by the presence of a strong band in Lane 1 around 300 bp. A faint band just above 700 bp is also visible, which sequencing demonstrated as off-target amplification. In Lanes 2 and 4, there is no evidence of AS of the glycosylation domain or presence of the cytoplasmic insert, respectively. The control reaction in Lane 3 targets the transmembrane region of the receptor, which demonstrates a diffuse band indicative of some nucleotide diversity. At the annotation level, an 18-nucleotide difference is annotated at the boundary of zebrafish exon 18, which appears to be unique to the species. We confirmed the presence of both +/− 18-nucleotide products in this band indicating the use of an alternative 3′ splice site.

In chickens, two bands were observed in Lane 1, just above 300 bp and below 200 bp, indicating the eighth LDLa repeat is alternatively spliced (Fig. 2*D*). A faint upper band between 400 and 600 bp is also apparent in Lane 1. Sequencing analysis indicated this band is a mixture of nonspecific amplification and sequence demonstrating inclusion of the eighth LDLa repeat. The band at 300 bp that appears like a doublet revealed only sequence indicating inclusion of the eighth LDLa repeat and no evidence of the 39-nucleotide exon encoding the furin insert, which is consistent with published literature [29,56]. To understand whether the exon encoding the furin cleavage site is present in the genome but constitutively spliced out at the RNA level, we examined the chicken *Lrp8* locus in comparison to the mouse *Lrp8* locus, which does contain this exon. We found no evidence of a conserved region in the chicken genome (Fig. S2), indicating the exon encoding the furin cleavage site was likely introduced into the genome at a point evolutionarily later than chickens. In Lane 2, the glycosylation domain appears to be constitutively included. Interestingly, in Lane 3 we detected an upper band around 300 bp in addition to the expected control product targeting the transmembrane region. Sequencing identified the upper band to be an intron retention event, capturing chicken *Apoer2* exon 17, the intervening intron and exon 18. We also saw no evidence of the cytoplasmic insert (Lane 4) in chickens as has been published [29].

Rabbits demonstrated more diversity at the RNA level than both zebrafish and chickens, as shown in Fig. 2*E*. In Lane 1, at least four distinct bands are present, indicating the AS of the eighth LDLa repeat as well as the furin insert, which are found in all four possible splicing combinations, including skipping of both exons in tandem. Sequencing of clones from the topmost diffuse band also returned a variation of rabbit *Apoer2* that has a unique exon following the eighth LDLa repeat, which maps to a region of the genome between the eighth LDLa repeat and the furin insert (Fig. S3). Lane 2 indicates constitutive inclusion of the glycosylation domain. Lane 4 confirms the introduction of the alternatively spliced cytoplasmic insert, which is dynamically spliced in and out. The presence of a third band is intriguing and suggests additional splicing diversity in this region. However, efforts to sequence the

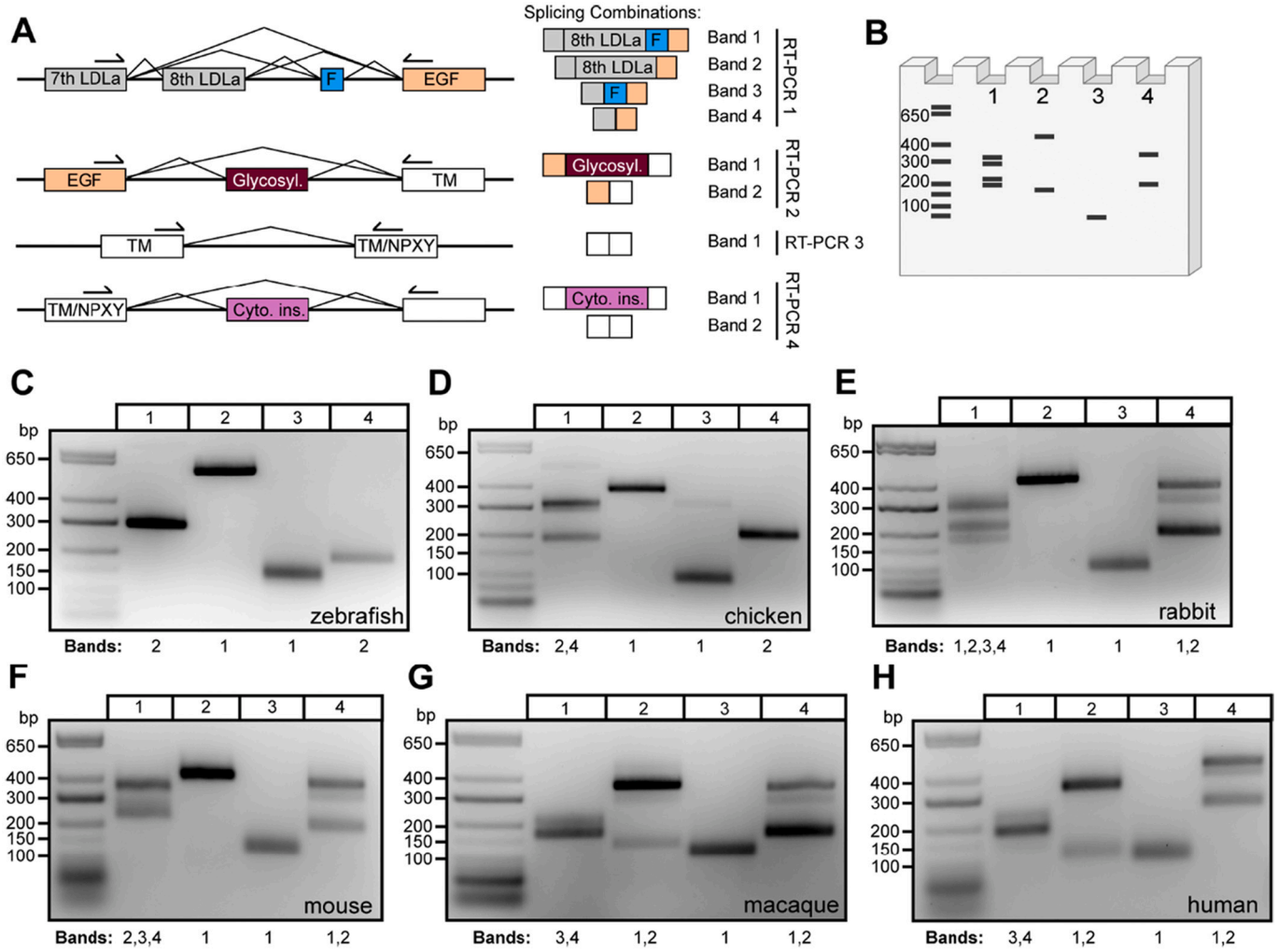


Fig. 2. *Apoer2* alternative splicing complexity in the brain increases across vertebrate evolution. (A) Schematic demonstrating RT-PCR primer design along with the corresponding exon to protein functional domain each primer set amplifies. Expected splicing combinations are depicted on the right side along with a band number corresponding to the cartoon agarose gel (B) (generated using Biorender.com) depicting potential RT-PCR banding patterns if all combinations of cassette exon decisions are used. (C–H) Gels depicting RT-PCR analysis of *Apoer2* alternative exons encoding the eighth LDLa repeat and furin insert (Lane 1), glycosylation domain (Lane 2), control transmembrane region (Lane 3) and cytoplasmic insert (Lane 4) across vertebrate species: (C) zebrafish, (D) chicken, (E) rabbit, (F) mouse, (G) macaque and (H) human. Bands in each lane are annotated below the gel according to the numbering displayed in (A).

third (middle) band have only returned sequence including or excluding the cytoplasmic insert, making it unclear exactly why there is an apparent nucleotide size difference between these bands. One sequenced TOPO clone returned sequence in which 67 bp at the end of rabbit exon 18 were excluded and 61 bp at the beginning of exon 19 were excluded, which would suggest the use of two alternative splice sites. As only one clone returned this sequence, it seems more likely to be a PCR or cloning artifact than AS events, although it is possible these are real AS events present at low frequency.

In the mouse cerebral cortex, multiple bands are apparent in Fig. 2*F* Lane 1, indicating AS of the eighth LDLa repeat and the furin insert. Sequencing of clones confirmed species containing the following combinations: eighth LDLa repeat but lacking the furin insert, lacking the eighth repeat but containing the furin insert and species lacking both exons. Interestingly, no clones were identified containing both the eighth LDLa repeat and the furin insert. In Lane 2, constitutive inclusion of the glycosylation domain was observed, despite published evidence suggesting this exon is alternatively spliced in mice [44]. After extensive optimization of RT-PCR parameters, including testing of multiple annealing temperatures, addition of PCR additives and use of an alternative polymerase, we failed to identify any bands indicating AS of the glycosylation domain in mice. Lane 4 indicates AS of the cytoplasmic insert, as well as the presence of a third band; similar to rabbits (Fig. 2*E* Lane 4), all sequenced clones returned mouse sequence either containing or lacking the entire cytoplasmic insert.

Loss of the eighth LDLa repeat becomes apparent in primates when examining the expression profile of the macaque. Fig. 2*G* Lane 1 demonstrates only two bands, neither of which contain the eighth LDLa repeat, but do contain AS of the 39-nucleotide furin insert. In Lane 2, AS of the glycosylation domain is observed, whereas this exon was constitutively included in all other lower vertebrates analyzed (Fig. 2*C–F*). Consistent with rabbits and mice, monkeys demonstrated AS of the cytoplasmic insert in Lane 4, as well as the presence of a third band shifted in size that returned sequence both including or lacking the entire cytoplasmic insert.

Lastly, we examined the AS pattern of *APOER2* in humans (Fig. 2*H*). As expected, we observed no inclusion of the eighth LDLa repeat but did observe exclusion and inclusion of the 39-nucleotide furin insert in Lane 1. Like macaques, AS of the glycosylation domain was observed in Lane 2. AS of the cytoplasmic insert is also conserved in humans, with bands observed indicating both its presence and absence. As seen in rabbits, mice and monkey, a third band is also present, for which sequencing

returned cassette splicing of the cytoplasmic insert.

Overall, the RT-PCR analysis indicates the addition of the furin insert and the cytoplasmic insert sometime between the evolution of chicken and rabbits and the loss of the eighth LDLa ligand binding repeat from mice to monkeys. Interestingly, we only observed AS of the glycosylation domain in primates, including monkeys and humans.

3.2. Full-length Apoer2 isoform mapping in mice and humans identifies homologous isoforms

In addition to AS of the exons encoding the eighth LDLa repeat, the furin insert, the glycosylation domain and the cytoplasmic insert, other *Apoer2* exons have been reported to be alternatively spliced in mouse and human through RT-PCR analyses [27,29,38–40]. Therefore, we asked whether these AS events are often found in different combinations or whether most AS events occur in isolation within the full-length context of the *Apoer2* transcript. To understand the full splicing landscape and complexity of *Apoer2* isoforms across the brain, we performed *Apoer2* specific long-read sequencing on RNA from the mouse and human cerebral cortices. Briefly, we performed *Apoer2* specific first strand cDNA synthesis on total RNA from the cerebral cortex of mice and humans as schematized in Fig. 3*A*. We then performed RT-PCR targeting the entirety of the coding sequence of *Apoer2*, with a forward primer binding at the ATG site in exon 1 and reverse primer binding at the TGA stop site in exons 19 and 20 for humans and mice, respectively. After size selection, nucleic acid purification and library preparation, *Apoer2* cDNA amplicons for mice and humans were subject to Pacific Bioscience's single-molecule real-time (SMRT) sequencing to generate full-length *Apoer2* transcript sequences.

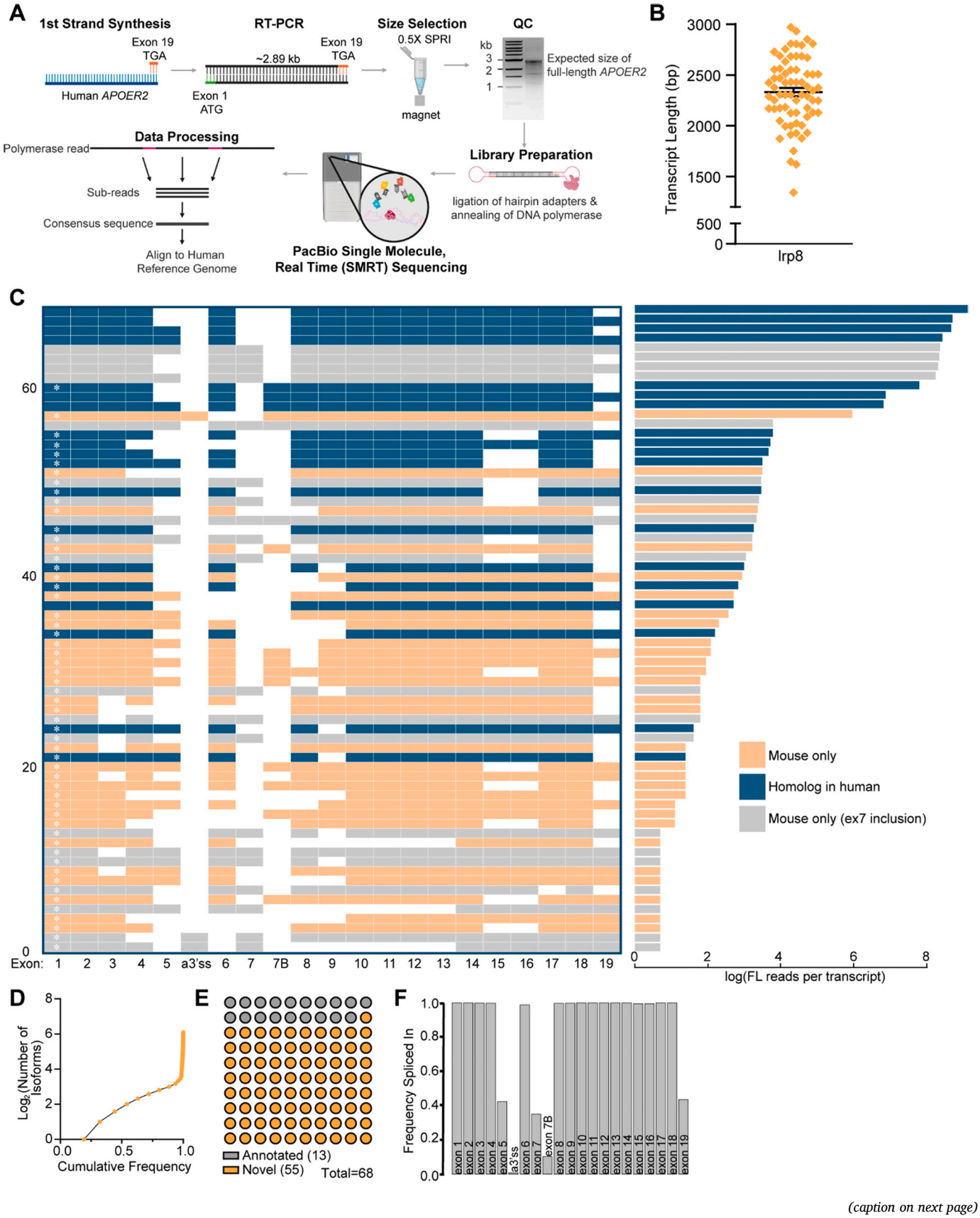
Sequencing data was processed into high quality isoforms then further collapsed into unique reference corrected isoforms. After data processing, the mouse and human samples yielded 3036 and 55,326 high quality full-length reads for analysis, with 87% and 84% respectively of those reads mapping to *Apoer2*. While the number of full-length, high-quality reads returned for the human was slightly lower than the mouse sample, the proportion of reads mapping to *Apoer2* was comparable, giving both RT-PCRs similar precision and enabling transcript mapping and proportional comparisons between the two species. For mouse *Apoer2*, we identified 68 high quality *Apoer2* transcripts that contained exons present in at least ten unique isoforms before filtering. As almost all introns have been spliced out, we are likely capturing mature processed RNA or genuine intron retention splicing events as opposed to pre-mRNA. The mean length of identified murine *Apoer2* transcripts clustered just under 2500 bp as shown in Fig. 3*B*. Full-length murine *Apoer2*, containing the coding sequence of all 21 exons captured within our primer design, would be about 3 kilobases. The comparatively lower average *Apoer2* transcript length suggests there are numerous AS events in the identified transcripts, with some transcripts likely having multiple AS events. When examining the 68 murine *Apoer2* transcripts, we see a plethora of cassette exon skipping events occurring individually and in tandem across the length of murine *Apoer2* (Fig. 3*C*). As our reverse primer bound at the nucleotides encoding the stop codon in the last coding exon (human exon 19 and mouse exon 20), we did not annotate this exon as a full exon in the transcript maps (Figs. 3*C* & 4*B*) since this full exon is much longer, encoding the 3'UTR which our RT-PCR scheme does not capture. The most abundant murine *Apoer2* isoform we identified contained three AS events, with exclusion of exons 5, 7 and 19, which encode 3 LDLa repeats, the eighth LDLa repeat and the cytoplasmic insert, respectively. Of interest, we identified novel AS of exon 3, which would remove an LDLa repeat, yet preserve the open reading frame, as well as transcripts that excluded exons 8 through 13 in tandem, which would remove the entire β-propeller domain and introduce a premature stop codon after the LDLa repeats. We also identified a known alternative 3′ splice site near exon 6, which adds 3 nucleotides of the prior intron, CAG, into the coding sequence of the transcript. Surprisingly, we identified multiple transcripts that contained tandem exon skipping of exons 15 and 16 which code for an EGF precursor-like repeat and the glycosylation domain respectively, while neither exon was identified to be skipped individually.

Of the 68 unique transcripts identified in the mouse cerebral cortex, majority of the AS events result in the inclusion or exclusion of entire protein functional domains at the protein level, suggesting AS may alter or regulate specific Apoer2 protein function in the brain. Based off the number of full-length reads per isoform (Fig. 3*C* right), we examined the cumulative frequency of individual *Apoer2* isoforms to understand how many isoforms make up the majority of *Apoer2* expression (Fig. 3*D*). We observed about 6–7 isoforms make up 75% of *Apoer2* expression in the mouse cerebral cortex, with the remaining isoforms found at lower levels. We also compared the 68 identified murine *Apoer2* isoforms to the coding region (excluding UTRs which our primer strategy does not capture) of those annotated in the Ensembl gene database [58]. Using this comparison and a manual confirmation and comparison to the NCBI [57] annotated transcripts as well, we identified 13 isoforms that have been previously annotated and 55 novel *Apoer2* isoforms (Fig. 3*C&E*).

Lastly, we wanted to understand the frequency at which individual exons are found to be spliced into the *Apoer2* transcript pool. We calculated a frequency spliced in value for each exon across all the isoforms weighted by their abundance as shown in Fig. 3*F*. Exon 5, encoding three LDLa repeats at the protein level, is spliced in only about 40% of the time, as is exon 7, encoding the eighth LDLa repeat, and exon 19, encoding the cytoplasmic insert. Exon 7B, encoding the furin insert, is included only about 15% of the time. While we observed many isoforms with AS of other exons, when these exons are examined on an individual level across all the transcripts weighted by their abundance, we see that most other exons are included at an almost constitutive level.

We also profiled human *APOER2* transcripts using a gene specific RT-PCR coupled with long-read sequencing and compared transcripts across the murine and human cerebral cortex (Figs. 3–4). We identified 48 high quality unique *APOER2* isoforms in the human cerebral cortex. These isoforms had a mean length just under 2500 bp (Fig. 4*A*), similar to the murine sample (Fig. 3*B*), likely indicating the presence of multiple alternative events since the expected length of the longest canonical *APOER2* product would be just over 2.9 kilobases. When we examine the individual human *APOER2* transcripts identified (Fig. 4*B*), the most abundant *APOER2* transcript is the canonical full-length *APOER2* transcript, closely followed by an *APOER2* transcript with exclusion of exon 18, encoding the cytoplasmic insert. Interestingly, in humans the third most abundant *APOER2* transcript contains cassette splicing of exon 15, encoding the glycosylation domain, which we did not observe to be alternatively spliced by itself in mice (Figs. 2*D* & 3*C*). We did observe tandem exclusion of exons 14 and 15, encoding an EGF precursor-like repeat and the glycosylation domains respectively, as was observed in mice (Fig. 3*C*). In humans, we observed one intron retention event, with the intron between exons 7 and 8 retained, and likely rendering that isoform a target for nonsense mediated decay. We also observed two alternative 3′ splice sites near exon 18, which encodes the cytoplasmic insert. Of these alternative 3′ sites, one was in the intron before exon 18, resulting in the inclusion of an additional 17 nucleotides, and the other was within exon 18, truncating the exon to only 61 base pairs. In mice, we did observe the presence of alternative 3′ splice sites at exon 19, corresponding to human exon 18; however, these transcripts dropped out of our exon abundance filtering. We analyzed whether those detected alternative sites in mouse were homologous to the two detected in human and did not observe homology at the protein level. However, the use of alternative splice sites at this exon boundary may be conserved.

We examined the cumulative frequency of the identified human *APOER2* transcripts (Fig. 4*C*) and found that 4–5 transcripts make up about 75% of *APOER2* expression in the human cerebral cortex, with the remaining 25% of expression made up of numerous unique isoforms. We also compared the identified isoforms to the coding regions of annotated *APOER2* transcripts in Ensembl [58] and NCBI [57] databases. Fourteen



(caption on next page)

Fig. 3. Full-length *Apoer2* isoform mapping in the murine cerebral cortex. (A) Schematic depicting *Apoer2* specific single molecule, long-read sequencing workflow (generated using Biorender.com). (B) Graph indicating mean transcript length ± SEM of unique murine *Apoer2* isoforms detected in the cerebral cortex excluding RT-PCR primers. (C) Left: Transcript matrix depicting *Apoer2* isoforms identified as individual rows. Transcripts with exons present in less than 10 unique transcripts before filtering were excluded, leaving 68 unique transcripts. Exons spliced in are colored, while skipped exons are white. Numbering at the left-hand margin indicates transcript number. A white asterisk inside of exon 1 indicates a novel or previously unannotated transcript. Right: Bar plot indicating the log of the total number of full-length reads of each corresponding transcript in the adjacent matrix. All transcripts and their corresponding number of reads are colored coded based on whether there is a corresponding homolog isoform in humans (blue), the transcript contains the mouse specific exon encoding the eighth LDLa repeat (grey), or whether the transcript is specific to mouse but does not contain the exon encoding the eighth LDLa repeat (orange). (D) Cumulative frequency of detected *Apoer2* isoforms. (E) Graph indicating the proportion of detected transcripts that are either annotated in NCBI or Ensembl or novel. (F) Bar graph demonstrating the frequency at which each *Apoer2* exon is spliced in based on the detected transcripts and their number of full-length reads. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

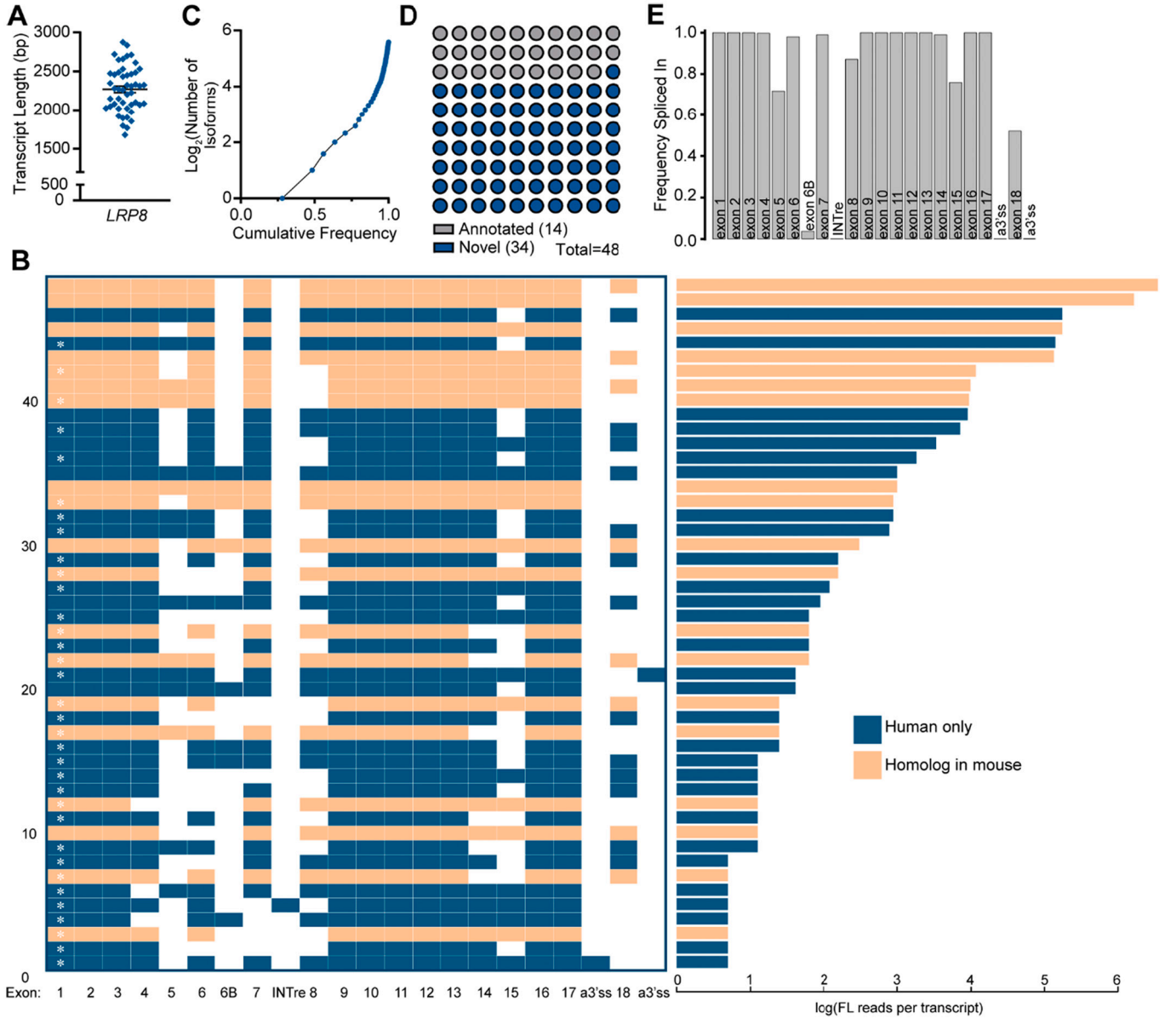


Fig. 4. *APOER2* full length isoform mapping in the human cerebral cortex. (A) Graph indicating the mean transcript length ± SEM of unique *APOER2* isoforms detected in the human cerebral cortex excluding RT-PCR primers. (B) Left: Transcript matrix displaying *APOER2* isoforms identified in the human cerebral cortex as individual rows. Exons spliced in are colored, while skipped exons are white. Numbering at the left-hand margin indicates transcript number. A white asterisk inside of exon 1 indicates a novel or previously unannotated transcript. Right: Bar graph displaying the log of the number of full-length reads per isoform in the adjacent transcript matrix. All transcripts and their corresponding number of reads are colored orange if there is a corresponding homolog isoform in mouse or blue if the isoform is specific to humans. (C) Cumulative frequency graph of detected *APOER2* isoforms. (D) Parts of a whole graph demonstrating the number of detected transcripts found to be annotated in NCBI or Ensembl databases or novel. (E) Bar graph indicating the frequency at which each *APOER2* exon is spliced in across all the detected isoforms weighted by their abundance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of the transcripts we identified are present in existing databases, whereas 34 of the isoforms described here are novel (Fig. 4*B&D*). Similar to the mouse, we calculated a frequency spliced in value for each individual exon across all the human isoforms (Fig. 4*E*). Interestingly, we find exon 5 is included at a higher frequency of about 70% in humans compared to mice (Fig. 3*F*). Exon 8, encoding an EGF precursor-like repeat in humans and corresponding to exon 9 in mice, is also spliced out more in humans, with a frequency spliced in value of about 85% compared to its frequency close to 100% in mice. Exon 15, encoding the glycosylation domain, is also dynamically spliced in humans, with a frequency spliced in of about 75%. Exon 18, encoding the cytoplasmic insert in humans, demonstrates a frequency spliced in value of about 50%, compared to the corresponding exon 19 in mice which is relatively similar around 40%.

3.3. Coordinated alternative splicing in mouse and human Apoer2

Since we observed that exons 15 and 16 were only spliced out in tandem in the mouse cerebral cortex (Fig. 3*C*), we wanted to examine the relationship of all individual AS events across *Apoer2* to determine

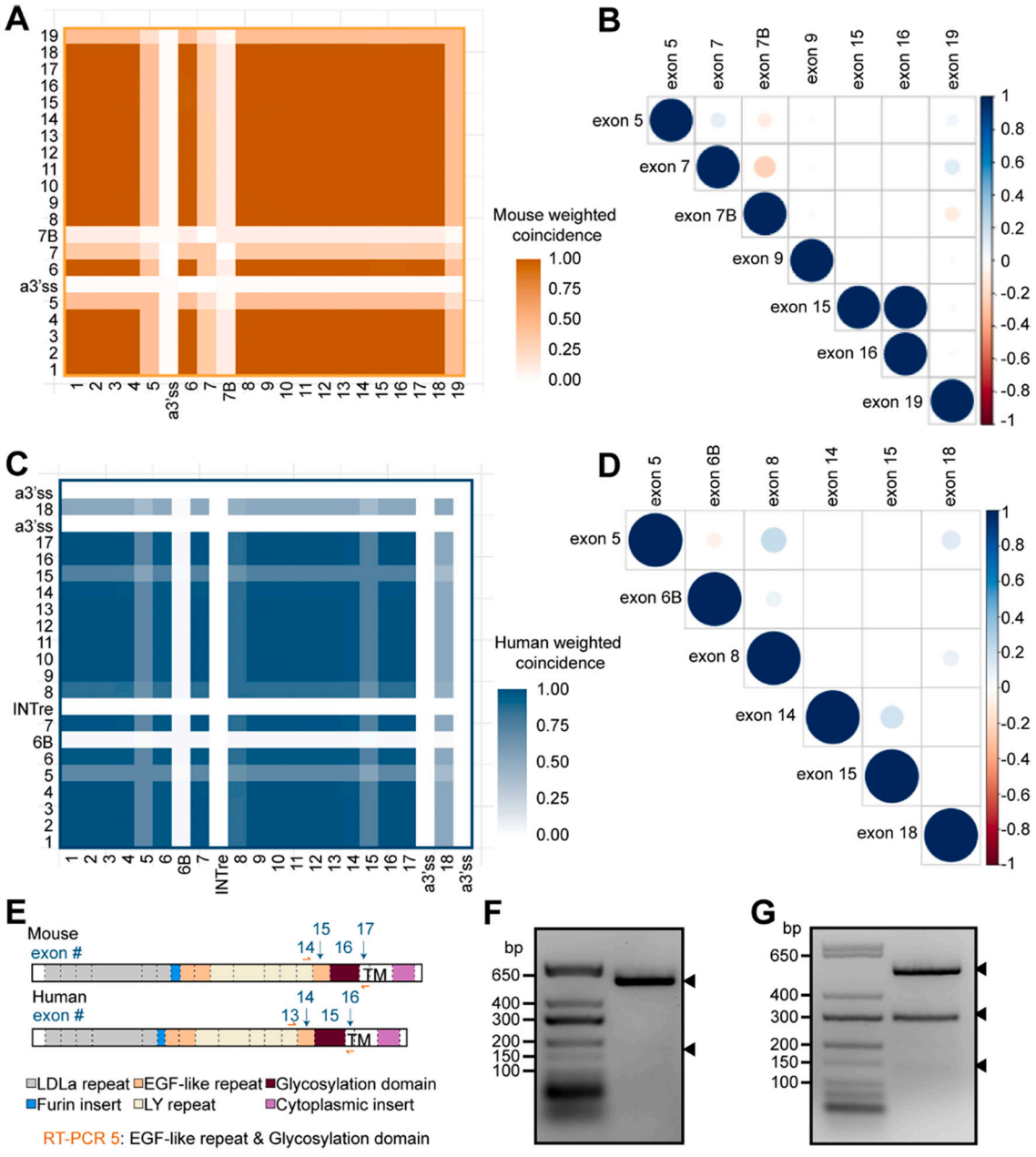


Fig. 5. Mouse *Apoer2* exhibits tandem splicing of exons encoding the third EGF precursor-like repeat and glycosylation domain. (A) Heatmap examining the coincidence of individual murine *Apoer2* exons compared to every other exon. Coincidence was calculated as the number of times the exons in each comparison pair were both spliced into the same transcript divided by the total number of transcripts. (B) Correlation matrix depicting Pearson's correlation coefficient of alternatively spliced murine *Apoer2* exons. Colour indicates value of Pearson's coefficient and size of dot indicates magnitude of significance, with a cutoff of $p < 0.01$. (C) Heatmap displaying coincidence value for human cerebral cortex *APOER2* exons. (D) Correlation matrix of Pearson's correlation coefficient analysis of human alternatively spliced *APOER2* exons. Significance is $p < 0.01$. (E) Schematic depicting mouse and human Apoer2 protein domains and corresponding coding exons along with RT-PCR strategy for F and G. (F) Gel depicting RT-PCR of mRNA from the murine cerebral cortex examining coordinated splicing of *Apoer2* exons 15 and 16. Arrowheads indicate detected bands that were excised and sequenced. (G) Gel analysis of RT-PCR from human cerebral cortex mRNA examining splicing of exons 14 and 15 in *APOER2*. Arrowheads indicate detected bands excised for sequencing confirmation.

whether splicing is coordinated across the length of the transcript. We calculated a weighted exon coincidence value, or the frequency at which each individual exon was spliced in at the same time as each other individual exon, (Fig. 5*A*, *C*). Exons such as 1–4 and 8–18, that are included at an almost constitutive level, exhibit high coincidence values as they are almost always spliced in together. Alternative cassette exons including 5, 6, 7, 7B and 19 demonstrate lower coincidence values due to their occasional exclusion. There also appears to be a shift in coincidence value between specific alternative exon pairs, such as exons 5 and 7, 5 and 7B, 7 and 7B, and exon 19 with 5, 7 and 7B. To examine these pairs more closely and determine whether there is any correlation between paired exon inclusion, we generated a correlation matrix (Fig. 5*B*). We also included exons 15 and 16 due to the tandem exclusion events we observed of these exons in Fig. 3*C*, as well as exon 9 for comparison with human exon 8 (Fig. 5*D*). Our results indicate a significant negative correlation between exon 7B and exons 5, 7 and 19, with the strongest correlation observed between exons 7 and 7B, which is supported by our RT-PCR data (Fig. 2*F*). We also observed significant positive correlations between exons 5 and 7, exons 5 and 19 and exon 7 and 19. Exons 15 and 16 demonstrate a perfect correlation of 1, which reflects the pattern of splicing observed for these exons previously (Fig. 3*C*).

When we examined coincidence across human *APOER2* exons (Fig. 5*C*), it again highlighted alternatively spliced cassette exons, which have lower coincidence values, including exons 5, 8, 15 and 18. Certain pairs of alternative exons demonstrated an apparent shift in coincidence values which may indicate coordinated splicing, such as the pairs exon 5 and 8, 5 and 18 and 8 and 18. Correlation analysis revealed significant positive correlation between exons 5 and 8, 5 and 18, 6B and 8, and 8 and 18 (Fig. 5*D*). A small but significant positive correlation is also observed between exons 14 and 15; however, this correlation is smaller when compared to the correlation observed between the corresponding murine exons 15 and 16 (Fig. 5*B*). We also observed a significant negative correlation between exon 5 and 6B in humans.

To validate the tandem splicing event that we observed in murine *Apoer2* of exons 15 and 16 compared to corresponding human *APOER2* exons 14 and 15, we designed RT-PCR primers spanning both exons (Fig. 5*E*) and performed RT-PCR on cDNA generated using oligo-dT primers. In mouse cerebral cortex, we observed two bands present (Fig. 5*F*), with the top band of stronger intensity. Upon sequencing, we confirmed the top band to contain both murine exons 15 and 16, and the bottom faint band to lack both exons. In humans, we identified the presence of 3 bands (Fig. 5*G*). Sequencing analysis confirmed the top band to contain exons 14 and 15 and the second band to contain exon 14 but lack exon 15. The smallest band contained tandem exclusion of both exons 14 and 15. We also identified one clone that contained exon 14 as well as a segment of 66 base pairs from the intron between exons 14 and 15, perhaps indicating another alternative cassette exon event (Fig. S4). Together, this confirmed that in mice the last EGF precursor-like repeat and the glycosylation domain are only tandemly excluded, whereas in humans the glycosylation domain can be spliced out individually or in combination with the preceding exon.

4. Discussion

In this study, we utilized a combination of RT-PCR and single molecule, long-read RNA sequencing to map the *Apoer2* isoform pool across vertebrates. We found clear evolution of the *Apoer2* gene over vertebrate development along with changes in AS patterns, indicating gene diversification at multiple levels. This suggests *Apoer2* has been specifically altered over time, likely to introduce new functional consequences at the protein level. Application of SMRT long-read sequencing to *Apoer2* has provided us with a new perspective on the complexity of AS events across individual *Apoer2* transcripts in the cerebral cortex of both mice and humans. We identified 68 and 48 unique *Apoer2* isoforms in the cerebral cortex of each mice and humans, respectively, with only 19 isoforms homologous between the two species. Our findings revealed that *Apoer2* AS events occur in a plethora of combinations across the entire transcript and differ between species. Our results parallel recent findings published by the Allen Brain Atlas Team that highlight gene expression divergence in homologous cell types between mouse and human cortex despite overall conserved cell types [59]. This diversification in gene expression patterns between mouse and humans emphasizes the importance of species-specific studies, particularly in humans when relating gene function to disease and human functional biology. The identified diversity in *Apoer2* isoform expression implicates differential functional effects at the protein level among isoforms and potentially between species and could influence ligand binding, receptor surface expression, receptor proteolysis, and downstream signaling events that affect synaptic function.

We have found that the Apoer2 exon encoding the eighth LDLa repeat is constitutively included in zebrafish and alternatively spliced in chickens, rabbits and mice, yet lost in primates, which is consistent with published literature [38,56]. The functional consequences of losing the eighth LDLa repeat in primates is unclear, but it is possible that loss of this eighth repeat changes the binding properties of primate Apoer2 compared to that in lower vertebrates, as Apoer2 binds numerous important ligands in addition to Apoe, such as Reelin [20,60], clusterin [61], selenoprotein P [62], and many others [63]. AS of the eighth LDLa repeat in combination with AS of the three LDLa repeats encoded by exon 5 has already been shown to decrease Apoer2 affinity for β-VLDL when compared to AS of exon 5 alone [39]. Furthermore, exclusion of the eighth LDLa repeat also increases the affinity of Apoer2 for Reelin fragments [32], suggesting perhaps loss of the eighth LDLa repeat occurred in primates to increase the affinity of Apoer2 for its endogenous ligands.

We have also shown that the 39-nucleotide exon encoding a furin cleavage site is present in rabbits, mice and primates, yet absent in zebrafish and chickens. Inclusion of this 39-nucleotide insert introduces a furin cleavage site at the protein level, which has been shown to be cleaved and to generate a soluble Apoer2 ectodomain fragment that acts as an antagonist to Reelin signaling [43], which is critical for neuronal migration in the developing brain [20] and ligand induced LTP in the adult brain [21]. It is unclear whether this secreted fragment would interfere with Apoer2-Apoe binding, although that seems likely. Furthermore, addition of the furin cleavage site between chickens and rabbits is especially interesting as chickens are not known to synthesize Apoe [41], suggesting perhaps diversification of the *Apoer2* locus occurred to help modulate Apoe and other ligand binding in higher vertebrates. There is precedence for altered splicing of transmembrane receptors having potent effects on signaling, as ephrin transmembrane receptors have been shown to have either repulsive or adhesive effects on cell migration depending on which isoform is expressed [64]. Furthermore, AS has already been shown to play a role in the Reelin signaling pathway in development, as suppression by the splicing factor Nova2 of a specific isoform of Disabled-1, the adaptor protein that binds to Apoer2, is required for proper neuronal migration [65].

Interestingly, we found that AS of the glycosylation domain alone in Apoer2 appears to be restricted to primates. In mice, exon 16, encoding the glycosylation domain, has previously been reported to be alternatively spliced in the brain [44]; however, despite thorough efforts to optimize RT-PCR conditions, we were not able to detect a cassette exon skipping event of solely exon 16 in mice. Intriguingly, in mice we did detect coordinated AS of exon 16 with exon 15, which encodes the third EGF precursor-like repeat in Apoer2. As exons 15 and 16 are in the same phase, the open reading frame is likely preserved; however, the functional consequences of this tandem skipping event are unclear. The glycosylation domain has been shown to regulate extracellular processing of Apoer2, as this region contains the extracellular cleavage site for matrix metalloproteases [34]. It has also been found that exclusion of exon 16 in combination with AS of the cytoplasmic insert has unique functional consequences in the mouse brain, most notably on Apoer2

cell surface levels, synaptic strength and long term potentiation [34]. It is possible that AS of the glycosylation domain, whether in tandem with the EGF precursor-like repeat before it or by itself, emerged in mice and primates as a regulatory mechanism of Apoer2 surface levels, which are critical for maintaining the proper balance of dendritic spines [35].

The cytoplasmic insert emerged somewhere between chickens and rabbits, likely with the divergence of placental mammals and marsupials, which have been shown to lack the cytoplasmic insert [56]. While it is unclear where the cytoplasmic insert originated, as it is unique and not present in other LDLR family members [40], it is clear that it adds functional complexity in Apoer2 biology. Inclusion of the cytoplasmic insert is necessary for Reelin-induced LTP [30], neuronal survival [31] and intracellular adaptor protein binding [66–68]. Inclusion of the cytoplasmic insert in APOER2 has also been positively correlated with human global cognition and shown to be lowered in the temporal cortex of individuals with Alzheimer's disease. Furthermore, increasing inclusion of the cytoplasmic insert in an amyloid mouse model of Alzheimer's disease with an antisense oligonucleotide can partially rescue some spatial learning deficits [42]. It is reasonable to posit that incorporation of the cytoplasmic insert into Apoer2 conferred isoform specific functions and helped shape higher level vertebrate regulation of learning and memory. Apoer2 also undergoes sequential cleavage at the membrane, first extracellularly by matrix metalloproteases, and then intramembranously by γ-secretase, releasing an intracellular domain (ICD) that translocates to the nucleus and activates an enhancer profile necessary for the transcription of learning and memory genes [69]. It remains to be determined which form of the Apoer2 ICD is actually transcriptionally active, as the cytoplasmic insert is located in the ICD and its inclusion or exclusion would naturally change the size and likely structure of the ICD.

From our long-read sequencing data, we observed 68 and 48 unique *Apoer2* isoforms in the cerebral cortex of each mice and humans, respectively. The AS events that make up the unique identified isoforms largely occur in regions that encode functional domains at the protein level. This suggests that the different combinations of these AS events will alter the properties of individual Apoer2 isoforms at the protein level. Since *Apoer2* AS events occur in many combinations, it highlights the need to understand how these events are regulated and whether they have diverse functional effects. Splicing is often controlled in a spatiotemporal manner [70], and so combinatorial splicing across the *Apoer2* transcript provides multiple points at which to fine tune Apoer2 function as needed at different times and places in the brain. Indeed, we observed coordinated splicing in murine *Apoer2*, where the third EGF precursor-like repeat and the glycosylation domain are only alternatively spliced out together. This suggests some coordinated action of splicing involving these two exons.

Our study utilized bulk tissue from the cerebral cortex for analysis in order to define a clear repertoire of diverse and novel full-length *Apoer2* isoforms and highlight species-specific differences in splicing decisions across the human and mouse cerebral cortex. It is likely that individual *Apoer2* isoforms are expressed in a spatiotemporal manner or in a cell-type specific manner [10], which could give biological impact to some of the lowly expressed transcripts identified. It is also possible that the biological function of having several unique isoforms present at low levels could be to detract from having expression of some of the more abundant isoforms, in a type of splicing regulatory system. Therefore, it will be important for future studies to determine the localization of some of these isoforms within the cerebral cortex as well as determine if they exhibit cell-type specificity or cell to cell heterogeneity. Examining transcript localization is no easy task due to the need to identify the specific combination of exons across the length of the entire transcript. However, new methods are being developed to examine full length transcripts within individual cell types [71,72], which could prove interesting in the case of *Apoer2* biology in the brain. Our finding of 68 unique *Apoer2* isoforms in the mouse cerebral cortex is not far off from the identified 50 *Apoer2* transcripts identified in the retina and brain by Ray and colleagues [73]. As such, we are confident our isoform profiling of *Apoer2* in mice and humans reflects true isoform diversity present in the brain.

Humans demonstrate substantial heterogeneity at the genomic level, that can affect RNA sequence and splicing decisions. Our study focused on defining the full repertoire of *Apoer2* isoforms in one sample per species to understand broadly the potential isoform diversity of *Apoer2*. We cannot exclude the fact that individual genetic variation will affect *Apoer2* isoform biology. Therefore, future studies utilizing biological replicates within species will be necessary to understand if and how human genetic diversity contributes to *APOER2* splicing diversity. With genomic sequencing of individual samples done in parallel, the field could also start to understand whether any genomic variation, perhaps in the form of single nucleotide polymorphisms, are associated with certain splicing choices, as has been seen with *CD33* [74] and *RBM23* [75].

Overall, we have demonstrated that the *Apoer2* gene has evolved over the vertebrate lineage from zebrafish to humans, both at the genomic and AS level. Particularly, *Apoer2* shows an extremely high number of AS events that result in the inclusion or exclusion of different functional domains in numerous combinations that implicate differential functions for Apoer2. Therefore, Apoer2 can then be thought of as a multi-functional receptor that can be tightly controlled depending on the AS pattern it possesses. Moving forward, it will be important to understand how the many AS events observed in this study contribute to modulating Apoer2's roles in brain development, synapse formation [35] and long-term potentiation [21] based on Apoe and Reelin binding [19,60].

Funding

This work was supported by the National Institutes of Health [grant numbers R01AG059762, F31AG069498, & 5T32GM008541-22]; and the Harold and Margaret Southerland Alzheimer's Research Fund.

Sequence data from this article has been deposited with the DDBJ/EMBL/GenBank Data Libraries under Submission ID 2549724.

CRediT authorship contribution statement

Christina M. Gallo: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Adam T. Labadorf:** Software, Writing – review & editing, Supervision. **Angela Ho:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Uwe Beffert:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

We would like to thank Dr. Shannon Fisher and lab for providing zebrafish brain tissue for this study. We also thank Dr. Rachel Flynn for critical feedback of experimental setup. We also thank Drs. Sara Goodwin and Ying Jin at the Cold Spring Harbor Sequencing Core for their assistance in the long-read sequencing PacBio experiment and initial SMRTLink data analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2022.110318.

References

[1] A. Frankish, M. Diekhans, I. Jungreis, J. Lagarde, J.E. Loveland, J.M. Mudge, C. Sisu, J.C. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, C. Boix, S. Carbonell Sala, F. Cunningham, T. Di Domenico, S. Donaldson, I.T. Fiddes, C. García Girón, J.M. Gonzalez, T. Grego, M. Hardy, T. Hourlier, K.L. Howe, T. Hunt, O.G. Izuogu, R. Johnson, F.J. Martin, L. Martínez, S. Mohanan, P. Muir, F. C.P. Navarro, A. Parker, B. Pei, F. Pozo, F.C. Riera, M. Ruffier, B.M. Schmitt, E. Stapleton, M.-M. Suner, I. Sycheva, B. Uszczynska-Ratajczak, M.Y. Wolf, J. Xu, Y. T. Yang, A. Yates, D. Zerbino, Y. Zhang, J.S. Choudhary, M. Gerstein, R. Guigó, T.J. P. Hubbard, M. Kellis, B. Paten, M.L. Tress, P. Flicek, GENCODE 2021, Nucleic Acids Res. 49 (2021) D916–D923, https://doi.org/10.1093/nar/gkaa1087.

[2] International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, Center for Genome Research, E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, Christina Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, The Sanger Centre, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, Washington University Genome Sequencing Center, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S. W. Clifton, US DOE Joint Genome Institute, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.-F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, Baylor College of Medicine Human Genome Sequencing Center, R.A. Gibbs, D.M. Muzny, S.E. Scherer, J.B. Bouck, E. J. Sodergren, K.C. Worley, C.M. Rives, J.H. Gorrell, M.L. Metzker, S.L. Naylor, R. S. Kucherlapati, D.L. Nelson, G.M. Weinstock, RIKEN Genomic Sciences Center, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, Genoscope and CNRS UMR-8030, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, Department of Genome Analysis, Institute of Molecular Biotechnology, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, GTC Sequencing Center, D.R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H.M. Lee, J. Dubois, Beijing Genomics Institute/Human Genome Center, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, Multimegabase Sequencing Center, The Institute for Systems Biology, L. Hood, L. Rowen, A. Madan, S. Qin, Stanford Genome Technology Center, R.W. Davis, N.A. Federspiel, A.P. Abola, M. J. Proctor, University of Oklahoma's Advanced Center for Genome Technology, B. A. Roe, F. Chen, H. Pan, Max Planck Institute for Molecular Genetics, J. Ramser, H. Lehrach, R. Reinhardt, Cold Spring Harbor Laboratory, Lita Annenberg Hazen Genome Center, W.R. McCombie, M. de la Bastide, N. Dedhia, GBF—German Research Centre for Biotechnology, H. Blöcker, K. Hornischer, G. Nordsiek, *Genome Analysis Group (listed in alphabetical order, also includes individuals listed under other headings), R. Agarwala, L. Aravind, J.A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D.G. Brown, C.B. Burge, L. Cerutti, H.-C. Chen, D. Church, M. Clamp, R.R. Copley, T. Doerks, S.R. Eddy, E.E. Eichler, T.S. Furey, J. Galagan, J.G.R. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L.S. Johnson, T.A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W.J. Kent, P. Kitts, E.V. Koonin, I. Korf, D. Kulp, D. Lancet, T.M. Lowe, A. McLysaght, T. Mikkelsen, J.V. Moran, N. Mulder, V.J. Pollara, C.P. Ponting, G. Schuler, J. Schultz, G. Slater, A.F.A. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y.I. Wolf, K. H. Wolfe, S.-P. Yang, R.-F. Yeh, Scientific management: National Human Genome Research Institute, US National Institutes of Health, F. Collins, M.S. Guyer, J. Peterson, A. Felsenfeld, K.A. Wetterstrand, Stanford Human Genome Center, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D.R. Cox, University of Washington Genome Center, M.V. Olson, R. Kaul, Christopher Raymond, Department of Molecular Biology, Keio University School of Medicine, N. Shimizu, K. Kawasaki, S. Minoshima, University of Texas Southwestern Medical Center at Dallas, G.A. Evans, M. Athanasiou, R. Schultz, Office of Science, US Department of Energy, A. Patrinos, The Wellcome Trust, M.J. Morgan, Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921, https://doi.org/10.1038/35057062.

[3] T.W. Nilsen, B.R. Graveley, Expansion of the eukaryotic proteome by alternative splicing, Nature 463 (2010) 457.

[4] E.T. Wang, R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S.F. Kingsmore, G.P. Schroth, C.B. Burge, Alternative isoform regulation in human tissue transcriptomes, Nature 456 (2008) 470–476, https://doi.org/10.1038/nature07509.

[5] T.A. Hughes, Regulation of gene expression by alternative untranslated regions, Trends Genet. 22 (2006) 119–122, https://doi.org/10.1016/j.tig.2006.01.001.

[6] Q. Li, J.-A. Lee, D.L. Black, Neuronal regulation of alternative pre-mRNA splicing, Nat. Rev. Neurosci. 8 (2007) 819–831, https://doi.org/10.1038/nrn2237.

[7] L. Cartegni, S.L. Chew, A.R. Krainer, Listening to silence and understanding nonsense: exonic mutations that affect splicing, Nat. Rev. Genet. 3 (2002) 285–298, https://doi.org/10.1038/nrg775.

[8] M. Krawczak, J. Reiss, DavidN Cooper, The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences, Hum. Genet. 90 (1992), https://doi.org/10.1007/BF00210743.

[9] A.V. Alekseyenko, N. Kim, C.J. Lee, Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes, RNA 13 (2007) 661–670, https://doi.org/10.1261/rna.325107.

[10] H. Keren, G. Lev-Maor, G. Ast, Alternative splicing and evolution: diversification, exon definition and function, Nat. Rev. Genet. 11 (2010) 345–355, https://doi.org/10.1038/nrg2776.

[11] E. Kim, A. Goren, G. Ast, Alternative splicing: current perspectives, Bioessays 30 (2008) 38–47, https://doi.org/10.1002/bies.20692.

[12] N.L. Barbosa-Morais, Systematic genome-wide annotation of spliceosomal proteins reveals differential gene family expansion, Genome Res. 16 (2005) 66–77, https://doi.org/10.1101/gr.3936206.

[13] J.A. Kolkman, W.P.C. Stemmer, Directed evolution of proteins by exon shuffling, Nat. Biotechnol. 19 (2001) 423–428, https://doi.org/10.1038/88084.

[14] M. Liu, A. Grigoriev, Protein domains correlate strongly with exons in multiple eukaryotic genomes – evidence of exon shuffling? Trends Genet. 20 (2004) 399–403, https://doi.org/10.1016/j.tig.2004.06.013.

[15] R.L. Dorit, W. Gilbert, The limited universe of exons, Curr. Opin. Genet. Dev. 1 (1991) 464–469, https://doi.org/10.1016/S0959-437X(05)80193-5.

[16] L. Patthy, Genome evolution and the evolution of exon-shuffling — a review, Gene 238 (1999) 103–114, https://doi.org/10.1016/S0378-1119(99)00228-0.

[17] T. Sudhof, D. Russell, J. Goldstein, M. Brown, R. Sanchez-Pescador, G. Bell, Cassette of eight exons shared by genes for LDL receptor and EGF precursor, Science 228 (1985) 893–895, https://doi.org/10.1126/science.3873704.

[18] G.W. Yeo, E. Van Nostrand, D. Holste, T. Poggio, C.B. Burge, Identification and analysis of alternative splicing events conserved in human and mouse, Proc. Natl. Acad. Sci. 102 (2005) 2850–2855, https://doi.org/10.1073/pnas.0409742102.

[19] D.H. Kim, H. Iijima, K. Goto, J. Sakai, H. Ishii, H.J. Kim, H. Suzuki, H. Kondo, S. Saeki, T. Yamamoto, Human apolipoprotein E receptor 2. A novel lipoprotein receptor of the low density lipoprotein receptor family predominantly expressed in brain, J. Biol. Chem. 271 (1996) 8373–8380, https://doi.org/10.1074/jbc.271.14.8373.

[20] M. Trommsdorff, M. Gotthardt, T. Hiesberger, J. Shelton, W. Stockinger, J. Nimpf, R.E. Hammer, J.A. Richardson, J. Herz, Reeler/disabled-like disruption of neuronal migration in knockout mice lacking the VLDL receptor and ApoE receptor 2, Cell 97 (1999) 689–701, https://doi.org/10.1016/S0092-8674(00)80782-5.

[21] E.J. Weeber, U. Beffert, C. Jones, J.M. Christian, E. Förster, J.D. Sweatt, J. Herz, Reelin and ApoE receptors cooperate to enhance hippocampal synaptic plasticity and learning, J. Biol. Chem. 277 (2002) 39944–39952, https://doi.org/10.1074/jbc.M205147200.

[22] M.E. Belloy, V. Napolioni, M.D. Greicius, A quarter century of APOE and Alzheimer's disease: progress to date and the path forward, Neuron 101 (2019) 820–838, https://doi.org/10.1016/j.neuron.2019.01.056.

[23] Y.-W.A. Huang, B. Zhou, M. Wernig, T.C. Südhof, ApoE2, ApoE3, and ApoE4 differentially stimulate APP transcription and Aβ secretion, Cell 168 (2017) 427–441.e21, https://doi.org/10.1016/j.cell.2016.12.044.

[24] Z. Qiu, B.T. Hyman, G.W. Rebeck, Apolipoprotein E receptors mediate neurite outgrowth through activation of p44/42 mitogen-activated protein kinase in primary neurons, J. Biol. Chem. 279 (2004) 34948–34956, https://doi.org/10.1074/jbc.M401055200.

[25] E.H. Corder, A.M. Saunders, W.J. Strittmatter, D.E. Schmechel, P.C. Gaskell, G. W. Small, A.D. Roses, J.L. Haines, M.A. Pericak-Vance, Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families, Science 261 (1993) 921–923.

[26] A.M. Saunders, W.J. Strittmatter, D. Schmechel, P.H. St. George-Hyslop, M. A. Pericak-Vance, S.H. Joo, B.L. Rosi, J.F. Gusella, D.R. Crapper-MacLachlan, M. J. Alberts, C. Hulette, B. Crain, D. Goldgaber, A.D. Roses, Association of apolipoprotein E allele 4 with late-onset familial and sporadic Alzheimer's disease, Neurology 43 (1993) 1467, https://doi.org/10.1212/WNL.43.8.1467.

[27] A.E. Clatworthy, W. Stockinger, R.H. Christie, W.J. Schneider, J. Nimpf, B. T. Hyman, G.W. Rebeck, Expression and alternate splicing of apolipoprotein E receptor 2 in brain, Neuroscience 90 (1999) 903–911, https://doi.org/10.1016/S0306-4522(98)00489-8.

[28] Y. Zhang, K. Chen, S.A. Sloan, M.L. Bennett, A.R. Scholze, S. O'Keeffe, H. P. Phatnani, P. Guarnieri, C. Caneda, N. Ruderisch, S. Deng, S.A. Liddelow, C. Zhang, R. Daneman, T. Maniatis, B.A. Barres, J.Q. Wu, An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex, J. Neurosci. 34 (2014) 11929–11947, https://doi.org/10.1523/JNEUROSCI.1860-14.2014.

[29] C. Brandes, S. Novak, W. Stockinger, J. Herz, W.J. Schneider, J. Nimpf, Avian and murine LR8B and human apolipoprotein E receptor 2: differentially spliced products from corresponding genes, Genomics 42 (1997) 185–191, https://doi.org/10.1006/geno.1997.4702.

[30] U. Beffert, E.J. Weeber, A. Durudas, S. Qiu, I. Masiulis, J.D. Sweatt, W.-P. Li, G. Adelmann, M. Frotscher, R.E. Hammer, J. Herz, Modulation of synaptic plasticity and memory by Reelin involves differential splicing of the lipoprotein receptor Apoer2, Neuron 47 (2005) 567–579, https://doi.org/10.1016/j.neuron.2005.07.007.

[31] U. Beffert, F. Nematollah Farsian, I. Masiulis, R.E. Hammer, S.O. Yoon, K.M. Giehl, J. Herz, ApoE receptor 2 controls neuronal survival in the adult brain, Curr. Biol. 16 (2006) 2446–2452, https://doi.org/10.1016/j.cub.2006.10.029.

[32] T. Hibi, M. Mizutani, A. Baba, M. Hattori, Splicing variations in the ligand-binding domain of ApoER2 results in functional differences in the binding properties to Reelin, Neurosci. Res. 63 (2009) 251–258, https://doi.org/10.1016/j.neures.2008.12.009.

[33] G. Rudenko, L. Henry, K. Henderson, K. Ichtchenko, M.S. Brown, J.L. Goldstein, J. Deisenhofer, Structure of the LDL receptor extracellular domain at endosomal pH, Science 298 (2002) 2353, https://doi.org/10.1126/science.1078124.
[34] C.R. Wasser, I. Masiulis, M.S. Durakoglugil, C. Lane-Donovan, X. Xian, U. Beffert, A. Agarwala, R.E. Hammer, J. Herz, Differential splicing and glycosylation of Apoer2 alters synaptic plasticity and fear learning, Sci. Signal. 7 (2014) ra113, https://doi.org/10.1126/scisignal.2005438.
[35] S.B. Dumanis, H.-J. Cha, J.M. Song, J.H. Trotter, M. Spitzer, J.-Y. Lee, E.J. Weeber, R.S. Turner, D.T.S. Pak, G.W. Rebeck, H.-S. Hoe, ApoE receptor 2 regulates synapse and dendritic spine formation, PLoS One 6 (2011), e17203, https://doi.org/10.1371/journal.pone.0017203.
[36] S.S. Minami, Y.M. Sung, S.B. Dumanis, S.H. Chi, M.P. Burns, E.-J. Ann, T. Suzuki, R. S. Turner, H.-S. Park, D.T.S. Pak, G.W. Rebeck, H.-S. Hoe, The cytoplasmic adaptor protein X11α and extracellular matrix protein Reelin regulate ApoE receptor 2 trafficking and cell movement, FASEB J. 24 (2010) 58–69, https://doi.org/10.1096/fj.09-138123.
[37] M. Trommsdorff, J.-P. Borg, B. Margolis, J. Herz, Interaction of cytosolic adaptor proteins with neuronal apolipoprotein E receptors and the amyloid precursor protein, J. Biol. Chem. 273 (1998) 33556–33560, https://doi.org/10.1074/jbc.273.50.33556.
[38] H.-J. Kim, D.-H. Kim, K. Magoori, S. Saeki, T.T. Yamamoto, Evolution of the apolipoprotein E receptor 2 gene by exon loss, J. Biochem. 124 (1998) 451–456, https://doi.org/10.1093/oxfordjournals.jbchem.a022134.
[39] C. Brandes, L. Kahr, W. Stockinger, T. Hiesberger, W.J. Schneider, J. Nimpf, Alternative splicing in the ligand binding domain of mouse ApoE receptor-2 produces receptor variants binding reelin but not alpha 2-macroglobulin, J. Biol. Chem. 276 (2001) 22160–22169, https://doi.org/10.1074/jbc.M102662200.
[40] D.-H. Kim, K. Magoori, T.R. Inoue, C.C. Mao, H.-J. Kim, H. Suzuki, T. Fujita, Y. Endo, S. Saeki, T.T. Yamamoto, Exon/intron organization, chromosome localization, alternative splicing, and transcription units of the human apolipoprotein E receptor 2 gene, J. Biol. Chem. 272 (1997) 8498–8504, https://doi.org/10.1074/jbc.272.13.8498.
[41] S. Novak, T. Hiesberger, W.J. Schneider, J. Nimpf, A new low density lipoprotein receptor homologue with 8 ligand binding repeats in brain of chicken and mouse, J. Biol. Chem. 271 (1996) 11732–11736, https://doi.org/10.1074/jbc.271.20.11732.
[42] A.J. Hinrich, F.M. Jodelka, J.L. Chang, D. Brutman, A.M. Bruno, C.A. Briggs, B. D. James, G.E. Stutzmann, D.A. Bennett, S.A. Miller, F. Rigo, R.A. Marr, M. L. Hastings, Therapeutic correction of ApoER2 splicing in Alzheimer's disease mice using antisense oligonucleotides, EMBO Molecular Med. 8 (2016) 328–345, https://doi.org/10.15252/emmm.201505846.
[43] S. Koch, V. Strasser, C. Hauser, D. Fasching, C. Brandes, T.M. Bajari, W. J. Schneider, J. Nimpf, A secreted soluble form of ApoE receptor 2 acts as a dominant-negative receptor and inhibits Reelin signaling, EMBO J. 21 (2002) 5996–6004, https://doi.org/10.1093/emboj/cdf599.
[44] P. May, H.H. Bock, J. Nimpf, J. Herz, Differential glycosylation regulates processing of lipoprotein receptors by γ-secretase, J. Biol. Chem. 278 (2003) 37386–37392, https://doi.org/10.1074/jbc.M305858200.
[45] J.S. Papadopoulos, R. Agarwala, COBALT: constraint-based alignment tool for multiple protein sequences, Bioinformatics 23 (2007) 1073–1079, https://doi.org/10.1093/bioinformatics/btm076.
[46] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, T.L. Madden, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, BMC Bioinform. 13 (2012) 134, https://doi.org/10.1186/1471-2105-13-134.
[47] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410, https://doi.org/10.1016/S0022-2836(05)80360-2.
[48] B. Treutlein, O. Gokce, S.R. Quake, T.C. Südhof, Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) E1291–E1299, https://doi.org/10.1073/pnas.1403244111.
[49] M. Tardaguila, L. de la Fuente, C. Marti, C. Pereira, F.J. Pardo-Palacios, H. del Risco, M. Ferrell, M. Mellado, M. Macchietto, K. Verheggen, M. Edelmann, I. Ezkurdia, J. Vazquez, M. Tress, A. Mortazavi, L. Martens, S. Rodriguez-Navarro, V. Moreno-Manzano, A. Conesa, SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification, Genome Res. 28 (2018) 396–411, https://doi.org/10.1101/gr.222976.117.
[50] R Core Team, R: A Language and Environment for Statistical Computing, 2020.
[51] J.T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, J.P. Mesirov, Integrative genomics viewer, Nat. Biotechnol. 29 (2011) 24–26, https://doi.org/10.1038/nbt.1754.
[52] H. Wickham, ggplot2: elegant graphics for data analysis, 2nd ed., Use R! Springer International Publishing: Imprint: Springer, Cham, 2016 https://doi.org/10.1007/978-3-319-24277-4.
[53] E. Flaherty, S. Zhu, N. Barretto, E. Cheng, P.J.M. Deans, M.B. Fernando, N. Schrode, N. Francoeur, A. Antoine, K. Alganem, M. Halpern, G. Deikus, H. Shah, M. Fitzgerald, I. Ladran, P. Gochman, J. Rapoport, N.M. Tsankova, R. McCullumsmith, G.E. Hoffman, R. Sebra, G. Fang, K.J. Brennand, Neuronal impact of patient-specific aberrant NRXN1α splicing, Nat. Genet. 51 (2019) 1679–1690, https://doi.org/10.1038/s41588-019-0539-z.
[54] Frank E. Harrell Jr., Charles Dupont and many others, Hmisc: Harrell Miscellaneous. R package version 4.4-2. https://CRAN.R-project.org/package=Hmisc, 2020.
[55] Taiyun Wei and Viliam Simko, R package "corrplot": Visualization of a Correlation Matrix (Version 0.84), Available from, https://github.com/taiyun/corrplot, 2017.
[56] N.B. Myant, Reelin and apolipoprotein E receptor 2 in the embryonic and mature brain: effects of an evolutionary change in the *apoER2* gene, Proc. R. Soc. B 277 (2010) 345–351, https://doi.org/10.1098/rspb.2009.1412.
[57] N.C.B.I. Resource Coordinators, R. Agarwala, T. Barrett, J. Beck, D.A. Benson, C. Bollin, E. Bolton, D. Bourexis, J.R. Brister, S.H. Bryant, K. Canese, M. Cavanaugh, C. Charowhas, K. Clark, I. Dondoshansky, M. Feolo, L. Fitzpatrick, K. Funk, L.Y. Geer, V. Gorelenkov, A. Graeff, W. Hlavina, B. Holmes, M. Johnson, B. Kattman, V. Khotomlianski, A. Kimchi, M. Kimelman, M. Kimura, P. Kitts, W. Klimke, A. Kotliarov, S. Krasnov, A. Kuznetsov, M.J. Landrum, D. Landsman, S. Lathrop, J.M. Lee, C. Leubsdorf, Z. Lu, T.L. Madden, A. Marchler-Bauer, A. Malheiro, P. Meric, I. Karsch-Mizrachi, A. Mnev, T. Murphy, R. Orris, J. Ostell, C. O'Sullivan, V. Palanigobu, A.R. Panchenko, L. Phan, B. Pierov, K.D. Pruitt, K. Rodarmer, E.W. Sayers, V. Schneider, C.L. Schoch, G.D. Schuler, S.T. Sherry, K. Siyan, A. Soboleva, V. Soussov, G. Starchenko, T.A. Tatusova, F. Thibaud-Nissen, K. Todorov, B.W. Trawick, D. Vakatov, M. Ward, E. Yaschenko, A. Zasypkin, K. Zbicz, Database resources of the National Center for Biotechnology Information, Nucleic Acids Res. 46 (2018) D8–D13, https://doi.org/10.1093/nar/gkx1095.
[58] K.L. Howe, P. Achuthan, James Allen, Jamie Allen, J. Alvarez-Jarreta, M.R. Amode, I.M. Armean, A.G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, M. Charkhchi, C. Cummins, L. Da Rin Fioretto, C. Davidson, K. Dodiya, B. El Houdaigui, R. Fatima, A. Gall, C. Garcia Giron, T. Grego, C. Guijarro-Clarke, L. Haggerty, A. Hemrom, T. Hourlier, O.G. Izuogu, T. Juettemann, V. Kaikala, M. Kay, I. Lavidas, T. Le, D. Lemos, J. Gonzalez Martinez, J.C. Marugán, T. Maurel, A. C. McMahon, S. Mohanan, B. Moore, M. Muffato, D.N. Oheh, D. Paraschas, A. Parker, A. Parton, I. Prosovetskaia, M.P. Sakthivel, A.I.A. Salam, B.M. Schmitt, H. Schuilenburg, D. Sheppard, E. Steed, M. Szpak, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, B. Walts, A. Winterbottom, M. Chakiachvili, A. Chaubal, N. De Silva, B. Flint, A. Frankish, S.E. Hunt, G.R. IIsley, N. Langridge, J. E. Loveland, F.J. Martin, J.M. Mudge, J. Morales, E. Perry, M. Ruffier, J. Tate, D. Thybert, S.J. Trevanion, F. Cunningham, A.D. Yates, D.R. Zerbino, P. Flicek, Ensembl 2021, Nucleic Acids Res. 49 (2021) D884–D891, https://doi.org/10.1093/nar/gkaa942.
[59] R.D. Hodge, T.E. Bakken, J.A. Miller, K.A. Smith, E.R. Barkan, L.T. Graybuck, J. L. Close, B. Long, N. Johansen, O. Penn, Z. Yao, J. Eggermont, T. Höllt, B.P. Levi, S. I. Shehata, B. Aevermann, A. Beller, D. Bertagnolli, K. Brouner, T. Casper, C. Cobbs, R. Dalley, N. Dee, S.-L. Ding, R.G. Ellenbogen, O. Fong, E. Garren, J. Goldy, R. P. Gwinn, D. Hirschstein, C.D. Keene, M. Keshk, A.L. Ko, K. Lathia, A. Mahfouz, Z. Maltzer, M. McGraw, T.N. Nguyen, J. Nyhus, J.G. Ojemann, A. Oldre, S. Parry, S. Reynolds, C. Rimorin, N.V. Shapovalova, S. Somasundaram, A. Szafer, E. R. Thomsen, M. Tieu, G. Quon, R.H. Scheuermann, R. Yuste, S.M. Sunkin, B. Lelieveldt, D. Feng, L. Ng, A. Bernard, M. Hawrylycz, J.W. Phillips, B. Tasic, H. Zeng, A.R. Jones, C. Koch, E.S. Lein, Conserved cell types with divergent features in human versus mouse cortex, Nature 573 (2019) 61–68, https://doi.org/10.1038/s41586-019-1506-7.
[60] G. D'Arcangelo, R. Homayouni, L. Keshvara, D.S. Rice, M. Sheldon, T. Curran, Reelin is a ligand for lipoprotein receptors, Neuron 24 (1999) 471–479.
[61] C. Leeb, C. Eresheim, J. Nimpf, Clusterin is a ligand for apolipoprotein E receptor 2 (ApoER2) and very low density lipoprotein receptor (VLDLR) and signals via the Reelin-signaling pathway, J. Biol. Chem. 289 (2014) 4161–4172, https://doi.org/10.1074/jbc.M113.529271.
[62] R.F. Burk, K.E. Hill, G.E. Olson, E.J. Weeber, A.K. Motley, V.P. Winfrey, L. M. Austin, Deletion of apolipoprotein E Receptor-2 in mice lowers brain selenium and causes severe neurological dysfunction and death when a low-selenium diet is fed, J. Neurosci. 27 (2007) 6207–6211, https://doi.org/10.1523/JNEUROSCI.1153-07.2007.
[63] P. Dlugosz, J. Nimpf, The Reelin receptors apolipoprotein E receptor 2 (ApoER2) and VLDL receptor, IJMS 19 (2018) 3090, https://doi.org/10.3390/ijms19103090.
[64] J. Holmberg, D.L. Clarke, J. Frisén, Regulation of repulsion versus adhesion by different splice forms of an Eph receptor, Nature 408 (2000) 203–206, https://doi.org/10.1038/35041577.
[65] M. Yano, Y. Hayakawa-Yano, A. Mele, R.B. Darnell, Nova2 regulates neuronal migration through an RNA switch in Disabled-1 signaling, Neuron 66 (2010) 848–858, https://doi.org/10.1016/j.neuron.2010.05.007.
[66] M. Gotthardt, M. Trommsdorff, M.F. Nevitt, J. Shelton, J.A. Richardson, W. Stockinger, J. Nimpf, J. Herz, Interactions of the low density lipoprotein receptor gene family with cytosolic adaptor and scaffold proteins suggest diverse biological functions in cellular communication and signal transduction, J. Biol. Chem. 275 (2000) 25616–25624, https://doi.org/10.1074/jbc.M000955200.
[67] X. He, K. Cooley, C.H.Y. Chung, N. Dashti, J. Tang, Apolipoprotein receptor 2 and X11/mediate apolipoprotein E-induced endocytosis of amyloid- precursor protein and -secretase, leading to amyloid- production, J. Neurosci. 27 (2007) 4052–4060, https://doi.org/10.1523/JNEUROSCI.3993-06.2007.
[68] W. Stockinger, C. Brandes, D. Fasching, M. Hermann, M. Gotthardt, J. Herz, W. J. Schneider, J. Nimpf, The Reelin receptor ApoER2 recruits JNK-interacting Proteins-1 and -2, J. Biol. Chem. 275 (2000) 25625–25632, https://doi.org/10.1074/jbc.M004119200.
[69] F. Telese, Q. Ma, P.M. Perez, D. Notani, S. Oh, W. Li, D. Comoletti, K.A. Ohgi, H. Taylor, M.G. Rosenfeld, LRP8-Reelin-regulated neuronal enhancer signature underlying learning and memory formation, Neuron 86 (2015) 696–710, https://doi.org/10.1016/j.neuron.2015.03.033.
[70] B. Raj, B.J. Blencowe, Alternative splicing in the mammalian nervous system: recent insights into mechanisms and functional roles, Neuron 87 (2015) 14–27, https://doi.org/10.1016/j.neuron.2015.05.004.
[71] J. Hård, J.E. Mold, J. Eisfeldt, C. Tellgren-Roth, S. Häggqvist, I. Bunikis, O. Contreras-Lopez, C.-S. Chin, C.-J. Rubin, L. Feuk, J. Michaëlsson, A. Ameur,

Long-read whole genome analysis of human single cells (preprint), Genomics. (2021), https://doi.org/10.1101/2021.04.13.439527.
[72] L. Tian, J.S. Jabbari, R. Thijssen, Q. Gouil, S.L. Amarasinghe, H. Kariyawasam, S. Su, X. Dong, C.W. Law, A. Lucattini, J.D. Chung, T. Naim, A. Chan, C.H. Ly, G. S. Lynch, J.G. Ryall, C.J.A. Anttila, H. Peng, M.A. Anderson, A.W. Roberts, D.C. S. Huang, M.B. Clark, M.E. Ritchie, Comprehensive characterization of single cell full-length isoforms in human and mouse with long-read sequencing (preprint), Genomics. (2020), https://doi.org/10.1101/2020.08.10.243543.
[73] T.A. Ray, K. Cochran, C. Kozlowski, J. Wang, G. Alexander, M.A. Cady, W. J. Spencer, P.A. Ruzycki, B.S. Clark, A. Laeremans, M.-X. He, X. Wang, E. Park, Y. Hao, A. Iannaccone, G. Hu, O. Fedrigo, N.P. Skiba, V.Y. Arshavsky, J.N. Kay, Comprehensive identification of mRNA isoforms reveals the diversity of neural cell-surface molecules with roles in retinal development and disease, Nat. Commun. 11 (2020) 3328, https://doi.org/10.1038/s41467-020-17009-7.
[74] P. van Bergeijk, U. Seneviratne, E. Aparicio-Prat, R. Stanton, S.A. Hasson, SRSF1 and PTBP1 are *trans* -acting factors that suppress the formation of a CD33 splicing isoform linked to Alzheimer's disease risk, Mol. Cell. Biol. 39 (2019), e00568-18, https://doi.org/10.1128/MCB.00568-18.
[75] D. Garrido-Martín, B. Borsari, M. Calvo, F. Reverter, R. Guigó, Identification and analysis of splicing quantitative trait loci across multiple tissues in the human genome, Nat. Commun. 12 (2021) 727, https://doi.org/10.1038/s41467-020-20578-2.