

Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Don Asafu-Adjaye

March 17, 2023

Identification of Tissue Specific Splice Variants and Isoforms of RNA Exosome in Mice

by

Don G. Asafu-Adjaye

Anita H. Corbett

Adviser

Biology

Anita H. Corbett

Adviser

Homa Ghalei

Committee Member

Roger Deal

Committee Member

2023

Identification of Tissue Specific Splice Variants and Isoforms of RNA Exosome in Mice

By

Don Asafu-Adjaye

Anita H. Corbett

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Biology

2023

Abstract

Identification of Tissue Specific Splice Variants and Isoforms of RNA Exosome in Mice

By Don Asafu-Adjaye

The RNA exosome is a 10-subunit complex required for the processing, turnover, and degradation of many classes of RNA. This complex is highly conserved, essential for survival, and ubiquitously expressed in all tissues and cell types. Nine of the ten subunits are structural. EXOSC1, EXOSC2, and EXOSC3 make up the cap of the complex. EXOSC4, EXOSC5, EXOSC6, EXOSC7, EXOSC8, and EXOSC9 make up the core. The cap and the core form a hollow barrel for which RNA is threaded through to reach the catalytic base subunit, DIS3, a 3'-5' endo/exoribonuclease. Although the RNA exosome is essential and expressed in all cell types, missense mutations in the genes encoding several structural subunits have been linked to rare diseases that primarily cause neurological pathology, many of them severe childhood neuronopathies. One explanation for why the RNA exosome subunit gene mutations give rise to brain-specific diseases is that the structural subunit genes could be differentially spliced in specific tissues. Little is understood about the tissue-specific expression of the RNA exosome subunit splice variants, despite the existence of known variants. Here, we examine the steady-state levels of RNA exosome subunit mRNA and protein across various tissues. We find that the RNA exosome subunit isoforms have differences that may influence the ability of the RNA exosome to interact with cofactors and/or substrates.

Identification of Tissue Specific Splice Variants and Isoforms of RNA Exosome in Mice

by

Don Asafu-Adjaye

Anita H. Corbett

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Biology

2023

Acknowledgements

I would like to express my sincere gratitude to the members of the Corbett lab for their advice and guidance during my time in the lab. I would especially like to thank Julia L. de Amorim, a PhD graduate student, for her invaluable supervision, encouragement, and support that have fostered my growth as a scientist. Her feedback allowed me to design and refine the results presented in my results and my thesis would not be possible without her guidance.

I would also like to acknowledge and give my warmest thanks to Dr. Anita Corbett for giving me the opportunity to work in her lab during my four years at Emory. I would also like to thank her for her advice, encouragement, and support during my years at Emory University.

Lastly, I want to thank my thesis committee members: Dr. Anita Corbett, Dr. Homa Ghalei, and Dr. Roger Deal, for their encouragement, insightful feedback, and suggestions.

Table of Contents

<i>Introduction</i>	1
<i>Materials and Methods</i>	3
<i>Results</i>	7
<i>Discussion</i>	12
<i>References</i>	18
<i>Figures</i>	20
<i>Supplemental figures</i>	28

Introduction

The RNA exosome is a 3'-5' exo/endoribonuclease complex critical for proper gene expression and post-transcriptional regulation, including the maturation of ribosomal RNA (rRNA) [1, 2]. The RNA exosome targets many classes of RNAs for processing, degradation, and turnover [3]. This complex comprises ten subunits: nine structural noncatalytic subunits and one enzymatic subunit (**Figure 1A**, PDB #6D6Q, **1B**). EXOSC1, EXOSC2, and EXOSC3 make up the cap of the complex and EXOSC4, EXOSC5, EXOSC6, EXOSC7, EXOSC8, and EXOSC9 make up the core. The cap and core form a barrel-shaped structure, through which RNA is threaded in a 3'-5' orientation to reach the catalytic subunit DIS3, at the base of the complex [4]. All RNA exosome subunits are highly conserved, essential, and ubiquitously expressed [5].

In the nucleus, the RNA exosome targets rRNAs, antisense RNAs, RNA-DNA hybrids (R-loops), and small noncoding RNAs (snRNAs) for processing and/or degradation [6-8]. In the cytoplasm, the RNA exosome targets mRNA for regulatory turnover, aberrant mRNAs for decay, and double stranded RNAs (dsRNAs) as a mechanism for viral defense [9-11]. The RNA exosome complex was initially identified in budding yeast in a screen for rRNA and is highly conserved in eukaryotes [2, 6, 12].

The RNA exosome is reported to target specific RNAs in either the nucleus or the cytoplasm via interactions with proteins termed cofactors. In the nucleus, the RNA exosome associates with cofactors including the DExH-box helicase MTR4, the TRAMP polyadenylation complex, and the MPP6 docking protein [13-15]. Cytoplasmic RNA exosome cofactors include the rRNA channeling SKI complex [16]. Given these critical

interactions with the RNA exosome complex, any changes that alter the composition or conformation of the complex could have consequences in the protein-protein interactions.

In fact, missense mutations in the RNA exosome subunit genes have been known to perturb interactions with cofactors. For example, mutations leading to single amino acid substitutions in cap subunit, *EXOSC3*, reduce the RNA exosome interactions with MPP6 [13]. The mutations in question are linked to pontocerebellar hypoplasia type 1B (PCH1B), a heritable and rare pediatric neurological disease characterized by cerebellar atrophy, progressive microcephaly, global development delay, and infantile spinal motor neuron disease [17]. There is currently no cure and those born with the disease typically do not survive adulthood. Since the initial link between *EXOSC3* mutations and PCH1B, a number of missense mutations in other structural subunits of the RNA exosome have also been linked to disease [18]. Multiple missense mutations in genes encoding structural subunits of the RNA exosome impact the cerebellum and cause a subtype of pontocerebellar hypoplasia (PCH), a group of autosomal recessive neurodegenerative disorders characterized by hypoplasia and progressive atrophy of the cerebellum and pons [19]. Patients with missense mutations in *EXOSC9* are diagnosed with PCH1D, which is characterized by cerebellar atrophy with spinal motor neuronopathy [20]. Patients with missense mutations in *EXOSC1* suffer from PCH1F, which is characterized by cerebellar hypoplasia, mega cisterna magna, mild dilatation of cerebellar folia, mild hypoplasia of corpus callosum, vermian hypoplasia, and mild delayed myelination [21]. Why mutations in genes that code for the ubiquitously expressed RNA exosome complex give rise to neurological diseases is not well understood.

One possible model that could explain tissue or cell-specific pathology of mutations in the RNA exosome subunit genes is that there could be perturbations in the interactions between the complex and important tissue-specific cofactors. Some of these interactions could be impacted if the protein isoforms of the structural subunits of the RNA exosome differ across cell types or tissues as is common in higher eukaryotes. However, alternative splicing of RNA exosome subunit transcripts has not yet been explored. We hypothesize there may be differences in alternative splicing in RNA exosome subunits in specific tissues. Here, we analyze the levels of subunit splice variants and protein isoforms in murine tissues as a mechanism to probe potential variant requirements in different tissue types. We find that mouse tissues have multiple splice variants of *Exosc3* and *Exosc1* but have one splice variant of *Exosc9*. There are different RNA and protein levels of *Exosc3*, *Exosc9*, and *Exosc1* splice variants and isoforms in tissues. One protein isoform of each subunit may be all that is required for proper functioning of the RNA exosome in the tissues examined. Structural differences between RNA exosome subunit isoforms may affect the ability of the RNA exosome to interact with RNA substrates and/or cofactors.

Materials and Methods

Genome Tissue Expression Project (GTEx)

GTEx Release v8 project includes whole genome sequencing (WGS) and RNA (RNA-seq) data from 54 tissues from 838 post-mortem individuals. Each genotyped tissue has at least 70 samples (GTEx consortium 2020) [22]. We analyzed the median transcripts per million (TPM) of *EXOSC3*, *EXOSC9*, and *EXOSC1* in the heart (left ventricle, n = 386; atrial appendage, n = 372), kidney (cortex, n = 73), and brain

(cerebellum, n = 209; cerebellar hemisphere, n = 175; frontal cortex, n = 175; cortex, n = 205).

RNA isolation and RT-PCR

RNA was isolated from murine tissues harvested from wild type male C57BL/6J mice (Jackson laboratory) using TRIzol reagent following the manufacturer's protocol (ThermoFisher Scientific). Tissues extracted (50-100 mg) include heart, kidney, cerebellum, and cortex. Reverse transcription (RT) was performed on three replicates of 2 mg of extracted RNA in 10 μ L reactions. cDNA was generated using a thermocycler program with the setting 42°C for 60 min, 95°C for 5 min, and 25°C hold. The resulting cDNA (500 ng) was amplified by PCR. Splice variants of murine *Exosc3*, *Exosc9*, and *Exosc1* gene sequences were identified using GenBank accession numbers from the National Center for Biotechnology Information (NCBI). Primer sequences are designed to capture splice variants of murine *Exosc3*, *Exosc9*, and *Exosc1* (IDT) (**Table S1**).

cDNA was amplified using a thermocycler PCR program with T_{anneal} of 52°C for *Exosc3* splice variant 1, 57°C for *Exosc3* splice variant 2, 56°C for *Exosc9*, and 54°C for *Exosc1*. *Exosc3* splice variant 1, *Exosc9*, and *Exosc1* were amplified with 30 PCR cycles. *Exosc3* splice variant 2 was amplified with 35 PCR cycles. The PCR products were resolved using a 2% agarose gel and ethidium bromide (New England Biolabs). Bands were visualized using a ChemiDoc Imaging System (Bio-Rad). Gel images were quantified using ImageJ v.153 software (NIH).

Protein purification and antibody production

The gene encoding the murine *Exosc3* open reading frame was cloned into pGEX-6P-2 plasmid (GE Healthcare Life Sciences (now Cytiva)) to create an N-terminally

glutathione-S-transferase (GST) tagged Exosc3. Recombinant GST-Exosc3 was expressed in *Escherichia coli* Rosetta 2(DE3). The GST fusion protein was purified by affinity chromatography on glutathione-Sepharose (GE Healthcare Life Sciences (now Cytiva)) in 300 mM NaCl, 50 mM HEPES/NaOH (pH7.5), 5% glycerol and 2 mM beta-mercaptoethanol (β ME). The fusion protein was eluted by addition of 30 mM reduced glutathione. The GST tag was removed using PreScission protease (GE Healthcare Life Sciences (now Cytiva)) and the Exosc3 protein was further purified by collecting the flowthrough of a second affinity chromatography on glutathione-Sepharose resin. The untagged protein was further purified on a Superdex200 size exclusion chromatography column equilibrated in 150 mM NaCl, 30 mM HEPES/NaOH (pH 7.5), 5% glycerol. Expression and purification of both recombinant GST-Exosc3 and untagged Exosc3 were confirmed by SDS-PAGE followed by Coomassie staining. The purified untagged protein was used as an immunogen to raise rabbit polyclonal antibodies by Josman, LLC. Sera containing anti-Exosc3 antibodies was collected 21 days after immunization and used directly for immunoblotting, immunoprecipitation, and immunofluorescence.

Immunoblotting

Mouse tissues were harvested and lysed on ice in RIPA buffer (19 mM Tris-HCl pH 8, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X 100, 0.1 % SDS, 0.1% sodium deoxycholate) containing protease inhibitors (Pierce protease mini tablets, A32953). Protein lysates were boiled in reducing sample buffer and resolved on 4-20% Criterion TGX polyacrylamide gels (Bio-Rad), then transferred onto a nitrocellulose membrane (0.2 μ m, Bio-Rad). The nitrocellulose membrane was washed three times for ten minutes each with blocking solution (1X TBST, 5% dry milk). Primary antibodies were detected using

species-specific horse radish peroxidase (HRP), conjugated mouse or rabbit secondary antibodies (Jackson ImmunoResearch) followed by incubation with enhanced chemiluminescence substrate (ECL, Sigma) or Amersham™ ECL™ Prime (VWR 89168-782). Chemiluminescence was detected by exposing blots using a ChemiDoc Imaging System (Bio-Rad). Immunoblots were quantified using ImageLab software compatible with the ChemiDoc Imaging System (Bio-Rad). Primary antibodies and dilutions as they appear: EXOSC3 (custom made; 1:3000), EXOSC9 (Bethyl A303-888A, 1:2000), EXOSC1 (Proteintech 12585-1-AP, 1:500), HSP90 (Santa Cruz sc-13119, 1:3000). All antibodies were validated prior to use by demonstrating that siRNA-mediated depletion of the protein detected led to loss of the band detected.

Statistical analysis

Comparisons between experimental groups were made using an unpaired student's t-test. All data are presented as means and standard error of the mean (SEM) (error bars). Asterisks (*) indicate statistical significance at P-value < 0.05.

In silico protein predictions

The Robetta RoseTTAFold [23] and PyMOL molecular visualization software v.2.5 (DeLano Scientific LLC; Palo Alto, California, USA) were used in structural modeling. Platforms were used with the cryo-EM structure (PDB #6D6Q) of the human nuclear RNA exosome at 3.45 Angstrom resolution [13]. Amino acid sequences identified by accession numbers of EXOSC3, EXOSC9, and EXOSC1 protein isoforms from NCBI were used to generate predicted protein structures (**Table S2**). The predicted isoforms and cryo-EM isoforms were aligned using PyMOL.

Results

To test the hypothesis that there could be different isoforms of the RNA exosome expressed in different cells or tissues, we focused on three structural subunits that have predicted alternative splicing based on information presented within the NCBI database. We focused on *EXOSC3*, *EXOSC1*, which both encode cap subunits, and *EXOSC9*, which encodes one of the core subunits. The domain structures of these structural RNA exosome subunits are shown in **Figure 1C**, which also depicts the predicted protein isoforms based on reported alternative splicing of these genes.

To begin exploring the steady state levels of the RNA exosome subunits across tissues, we analyzed bulk gene expression by exploiting the Genotype Tissue Expression (GTEx) database. The GTEx database combines hundreds of RNA sequencing results from tissues of post-mortem humans. Steady state expression levels were examined by median transcript per million (TPM) mapped reads (**Figure 2**). These data collect all transcripts that map back to the corresponding genes, which include multiple splice variants of the same gene. Violin plots of bulk transcripts of *EXOSC3* (**Figure 2A**), *EXOSC9* (**Figure 2B**), and *EXOSC1* (**Figure 2C**) are shown for heart (left ventricle and atrial appendage), kidney, and brain (cerebellum, cerebellar hemisphere, frontal cortex, and cortex).

For *EXOSC3*, the cerebellum (TPM = 9.148) and the cerebellar hemisphere (TPM = 9.693) have higher median transcript levels when compared to the heart (left ventricle) (TPM = 2.676), heart (atrial appendage) (TPM = 4.036), kidney (TPM = 5.529), brain frontal cortex (TPM = 6.073), or brain cortex (TPM = 4.803) (**Figure 2A**). The transcripts

for *EXOSC9* across tissues have an even more notable difference (**Figure 2B**). Similar to *EXOSC3*, *EXOSC9* transcript levels are highest in the cerebellum (TPM = 18.95) and cerebellar hemisphere (TPM = 23.36). The heart (left ventricle, TPM = 4.524), heart (atrial appendage, TPM = 5.831), kidney (TPM = 5.304), brain frontal cortex (TPM = 8.336), and brain cortex (TPM = 6.640) have lower median transcript levels compared to the cerebellum and cerebellar hemisphere. Overall, *EXOSC9* transcript levels in the cerebellum and cerebellar hemisphere are higher compared with *EXOSC3* transcript levels. The transcript levels for *EXOSC1* are again highest in the cerebellum (TPM = 21.52) and cerebellar hemisphere (TPM = 24.58) compared with the heart (left ventricle, TPM = 9.351), heart (atrial appendage, TPM = 12.84), kidney (TPM = 15.86), brain frontal cortex (TPM = 20.51), and brain cortex (TPM = 17.47) (**Figure 2C**). Overall, the transcript levels of *EXOSC1* are highest across tissues compared with the transcript levels of *EXOSC9* and *EXOSC3*. These results show that the cerebellum and cerebellar hemisphere have the highest transcript levels of the RNA exosome subunits examined compared with other tissues. The results also indicate that *EXOSC9* and *EXOSC1* have higher median steady state transcript levels in cerebellum and cerebellar hemisphere compared to *EXOSC3*.

The data collected from GTEx compares all transcript splice variants of a single gene. We hypothesize that there may be a difference in the levels of splice variants of RNA exosome subunit genes across tissues. To explore alternate splice variants of the RNA exosome subunits genes for *Exosc3*, *Exosc9*, and *Exosc1* in tissues, we extracted total RNA from murine heart, kidney, cerebellum, and cortex, and used Reverse Transcriptase Polymerase Chain Reactions (RT-PCR) to amplify specific splice

variants. Splice variants 1 of *EXOSC3*, *EXOSC9*, and *EXOSC1* are the canonical splice variants.

In the tissues examined, two splice variants of *Exosc3* (**Figure 3A**) are detected using RT-PCR (**Figure 3B**) and quantified (**Figure 3C**). The asterisk (*) in Figure 3B indicates a non-specific band. Steady state splice variant 1 levels are higher than steady state splice variant 2 levels in heart, kidney, cerebellum, and cortex tissues (**Figure 3C**). Steady state RNA levels of the *Exosc3* splice variant 1 are significantly higher than the levels of *Exosc3* splice variant 2 in the cerebellum.

Exosc9 has one splice variant in murine tissues according to NCBI. We detect the single splice variant of *Exosc9* (**Figure 3D**) in heart, kidney, cerebellum, and cortex in mice (**Figure 3E**). Quantification of the RT-PCR results shows statistically significant differences in the levels of *Exosc9* in the cerebellum in comparison with the kidney (**Figure 3F**).

Primers are designed to capture the five splice variants of *Exosc1* (**Figure 3G**). We captured detectable levels of four of the five splice variants of *Exosc1* in the tissues analyzed (**Figure 3H**). Splice variant 1 of *Exosc1* is most abundant compared with the other splice variants in all tissues analyzed (**Figure 3H, Figure 3I**). There is significantly more *Exosc1* splice variant 1 in the cerebellum than splice variant 2, splice variant 3, splice variant 4, or splice variant 5.

We analyzed transcripts across mouse tissues; however, transcript levels do not always correlate with protein levels [24]. To compare the analysis of alternatively spliced transcripts with protein levels, we performed immunoblots to detect *Exosc3*, *Exosc9*, and *Exosc1* across tissues. The antibodies employed for these analyses were validated using

siRNA depletion prior to use. Results of these immunoblotting analyses are presented in **Figure 4**. Isoform 1 of Exosc3, Exosc9, and Exosc1 are considered the canonical isoforms. Results show one band corresponding to the molecular weight of Exosc3 isoform 1 in heart, kidney, cerebellum, and cortex (**Figure 4A**). The asterisk indicates non-specific bands. Exosc3 isoform 2 is not detected in the tissues examined. Exosc3 isoform 1 levels were higher in the kidney compared to the heart, cerebellum, and cortex. The cortex shows the lowest protein levels of Exosc3 isoform 1 compared with the other tissues. We normalized Exosc3 isoform 1 steady state levels to the cerebellum and detect more Exosc3 isoform 1 in the kidney when compared to the heart, cerebellum, and cortex (**Figure 4B**). The immunoblots reveal a band at the predicted molecular weight for Exosc9 of ~70 kDa in the cerebellum, and cortex for a single isoform (**Figure 4A**). The immunoblot analysis detected one isoform of Exosc1 in the heart, kidney, cerebellum, and cortex tissues at the predicted molecular weight of ~35 kDa (**Figure 4A**). We detect more steady state Exosc1 protein levels in the kidney compared to the heart, cerebellum, or cortex (**Figure 4B**). Overall, these results show that one isoform of Exosc3, Exosc9, and Exosc1 is most abundant in the tissues analyzed; however, the levels of these RNA exosome complex subunits vary across tissues.

Numerous reasons explain why we were unable to detect multiple protein isoforms of the subunits examined. However, we predict a protein product for the noncanonical isoforms may exist because the splice variants that code for the isoforms have start and stop codons within open reading frames. We therefore sought to analyze how the structural differences conferred by alternative splicing to produce distinct human protein isoforms of EXOSC3, EXOSC9, and EXOSC1 could influence the RNA exosome

structure and, potentially, function. The amino acid sequences of EXOSC3, EXOSC9, and EXOSC1 identified through accession numbers (**Table S2**) from NCBI were used to generate predicted isoform proteins through Robetta RoseTTAFold modeling method. In this analysis, we align the predicted protein structures to the cryogenic electron microscopy (cryo-EM) generated human RNA exosome subunit structures (PDB #6D6Q) with the PyMOL molecular visualization system (see Materials and Methods). We designate all protein isoforms in the cryo-EM-generated structure as isoform 1.

The alignment comparing the cryo-EM-generated EXOSC3 isoform 1 structure and the predicted structure EXOSC3 isoform 2 shows differences in the incorporation of the RNA exosome complex (**Figure 5A**). Part of the S1 motif and RNA binding KH domain (see **Figure 1C**) represented by the circled region is absent in isoform 2 (**Figure 5A**). Two isoforms of human EXOSC9 amino acid sequence are identified through accession numbers on NCBI (**Table S2**). The cryo-EM generated EXOSC9 isoform 1 structure failed to model 150 amino acids (293-443) that are present in the EXOSC9 isoform 1 amino acid sequence [13]. We therefore used RoseTTAFold to generate a predicted structure of EXOSC9 isoform 1 in order to obtain a predicted structure for the unmodeled amino acids in the cryo-EM-generated structure. We aligned the predicted isoform 1 structure to the cryo-EM generated isoform 1 structure (**Figure 5B.i**). The predicted isoform 1 structure model shows a region that is not present in the cryo-EM-generated structure. Apart from this difference, there are no other notable differences in the alignment of the two structures. EXOSC9 isoform 2 lacks 17 amino acids (385-402) present in isoform 1. Because these amino acids absent in EXOSC9 isoform 2 are not modeled in the cryo-EM-generated structure, we aligned the predicted isoform 2 structure to the predicted

isoform 1 structure (**Figure 5B.ii**). Results of this analysis show that the predicted EXOSC9 isoform 1 and predicted isoform 2 structures are different in the orientation and structure of the region missing in the cryo-EM-generated structure. The predicted structures of EXOSC1 isoform 2, isoform 3, isoform 4, and isoform 5 all lack sequences within the S1 binding motif that is present in the EXOSC1 isoform 1 structure (see **Figure 1C**). This difference in amino acid sequence between the isoforms leads to changes in secondary structure. The predicted structures of EXOSC1 isoform 2, isoform 3, isoform 4, and isoform 5 all lack the beta sheet clusters that are present in the cryo-EM-generated isoform 1 structure (**Figures 5C.i, 5C.ii, 5C.iii, 5C.iv**). The predicted isoform 5 structure forms an alpha helix structure that was not present in the cryo-EM-generated EXOSC1 structure. Overall, there are notable and potentially impactful differences between the isoforms of the subunits examined that could play a biologically relevant role in RNA exosome complex formation and interactions.

Discussion

This study characterizes the differences in steady state levels of RNA exosome subunits, *EXOSC3*, *EXOSC9*, and *EXOSC1* bulk transcript levels, splice variants, and isoforms in the heart, kidney, cerebellum, and cortex. We hypothesized that some tissues may require more RNA exosome levels than others or that alternative splicing could lead to heterogeneity of the RNA exosome in different tissues. The study validated a number of different splice variants for two structural subunits of the RNA exosome; however, a single predominant protein isoform was present in all tissues analyzed. The study revealed differences in the amount of transcript relative to protein, which could suggest

significant regulation of the levels of RNA exosome subunits at the protein level. This is the first study to address possible tissue-specific differences in the RNA exosome, despite missense mutations in RNA exosome genes cause pathology particularly in certain regions of the brain.

Initial analysis uses existing GTEx data. This GTEx data reveals that the cerebellum and cerebellar hemisphere maintained the highest levels of *EXOSC3*, *EXOSC9*, and *EXOSC1* compared to the heart (atrial appendage and left ventricle), kidney, frontal cortex, and cortex tissues (**Figure 2**). These results may suggest that the cerebellum and cerebellar hemisphere requires more RNA exosome-mediated RNA processing and/or turnover compared with the other tissues examined. These results may also explain why exosomopathy patients with missense mutations in structural RNA exosome subunit genes have the most well-defined phenotypes in the cerebellum compared with other tissues.

On the basis of the GTEx results, we expected RT-PCR to show the highest steady state RNA levels of *Exosc3*, *Exosc9*, and *Exosc1* in mouse cerebellum; however, results suggest that only the *Exosc9* maintains high steady state RNA levels in the cerebellum (**Figure 3**). These differences between the RT-PCR results and the GTEx data may be a result of the inherent differences in the techniques used to measure transcripts. The GTEx database measures the transcript levels of all splice variants of *EXOSC3*, *EXOSC9*, and *EXOSC1* in humans using RNA sequencing [22], while the RT-PCR experiments are designed to only detect specific splice variants of mouse RNA exosome subunits. Therefore, the inclusion of all splice variants in the GTEx genotyping of corresponding genes may have contributed to the differences in the transcript levels between the GTEx

and RT-PCR data. Additionally, RT-PCR is semi-quantitative while RNA sequencing is quantitative. Furthermore, the GTEx data are compiled from sequencing human tissues and we employed mouse tissues for analysis, and therefore species-specific differences likely exist. In addition, variables, such as age or sex which were not considered here, could influence the expression of RNA exosome subunits. In future, exploring transcript variants and protein isoforms in human tissue in age- and sex-specified groups will resolve these concerns.

Findings using GTEx show that there is higher steady state transcript levels of *EXOSC9* and *EXOSC1* in all tissues when compared to the transcript levels of *EXOSC3* in tissues (**Figure 2**). This finding was surprising because RNA exosome subunits are not linked to individual functions independent from the well-defined roles within the RNA exosome complex. The high transcript levels of in *EXOSC1* especially, suggest that this subunit is required independently of the RNA exosome complex. However, an increase in transcript levels may not translate to a change in protein levels as there are many levels of additional regulation [25].

Multiple splice variants of *Exosc3* and *Exosc1* were detected in different tissues. Despite detecting splice variants for these two transcripts, splice variant 1 for both *Exosc3* and *Exosc1* was more abundant than other splice variants in all tissues (**Figure 3**). These results suggest that splice variant 1 of *Exosc3* and *Exosc1* may encode the predominant subunits present in the RNA exosome complex as compared to the other splice variants for *Exosc3* and *Exosc1*. Consistent with this model, there are detectable steady state protein levels of one isoform for both *Exosc3* and *Exosc1* (**Figure 4**). The level of *Exosc3* isoform 1 was low in the cerebellum and cortex as compared to other tissues analyzed;

however, *Exosc3* splice variant 1 was not significantly different across all tissues analyzed, which suggests that there may be post-transcriptional, translational, and post-translational regulation that influences differences between steady state RNA and protein levels of *Exosc3*. Conversely, the protein levels of the one splice variant of *Exosc9* correlate with the protein levels of *Exosc9* in tissues (**Figure 3F, 4B**) suggesting that different regulatory mechanisms are involved in determining the steady state RNA and protein levels for *Exosc1* and *Exosc3* compared to *Exosc9*.

Based on the data obtained for transcripts, we expected to find multiple protein isoforms of *Exosc3* and *Exosc1* in tissues (**Figure 3**). In contrast to the expectation, we did not identify any abundant forms of *Exosc3* or *Exosc1* beyond the canonical isoform 1. This suggests that the noncanonical splice variants may not be translated, may be translated at undetectable levels, or the proteins produced are unstable. All the splice variants for *Exosc3* and *Exosc1* contain a start codon for translation; however, splice variant 2 of *Exosc3* lacks part of the 5' region present in the *Exosc3* splice variant 1 (**Figure 3A**). As a result, *Exosc3* splice variant 2 may not contain the proper ribosomal binding site required for translation, which may mean the transcript is translated at undetectable levels or not at all. Possibly, the isoforms of *Exosc1* and *Exosc3* encoded by transcript variants may be quickly degraded because the isoforms do not form the necessary secondary and tertiary structures needed for proper assembly into the RNA exosome complex. The human EXOSC3 and EXOSC1 noncanonical isoforms lack alpha helix and beta strand secondary structures corresponding to S1-like and S1 motif that are present in isoform 1 of EXOSC3 and EXOSC1 (**Figure 1C, Figure 5**). The lack of these secondary structures could possibly interfere with the function of the RNA exosome which

would cause these proteins to be degraded quickly to prevent possible interference with RNA exosome structure and/or function. For example, human EXOSC3 isoform 2 lacks alpha helices that form a portion of the S1 and KH binding motif, a conserved region of EXOSC3 that interact with RNA exosome cofactor MPP6 which is important for pre-rRNA and pre-mRNA surveillance and degradation (**Figure 1C, Figure 5A**) [26, 27]. Similarly, differences between the human EXOSC1 isoform 1 and noncanonical isoforms lead to a partial loss of the S1 motif, an important RNA interaction site [28], which could result in loss of interactions of noncanonical isoforms of EXOSC1 with RNA substrates. One approach to distinguish whether various protein isoforms are not produced or are rapidly degraded would be to introduce protease inhibitors in immunoblotting studies. Such an approach could provide insight into how the assembly of the RNA exosome complex is monitored.

This work presents the first attempt to analyze the steady state levels of RNA exosome subunit splice variants and isoforms comparing different tissues. The results of this work provide insight into potential regulatory mechanisms, and the relative abundance of specific RNA exosome splice variants and isoforms in different tissues. Using the relative abundance of specific RNA exosome subunits in tissues, we present evidence to suggest that certain tissues may require higher levels of subunit splice variants or isoforms of specific RNA exosome subunits for proper function. Furthermore, this work provides a basis for understanding how the different isoforms of EXOSC3, EXOSC9, and EXOSC1 may influence the structure of the RNA exosome and how the differences in isoforms may affect the ability of the RNA exosome to bind with cofactors or RNA substrates. Further analysis of RNA exosome subunits in specific cell types, such

as those that are abundant in the cerebellum, could provide insight into why mutations in genes encoding structural subunits of the RNA exosome cause specific pathology.

Acknowledgements: The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on March 1, 2023.

References

1. Mitchell, P., et al., *The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonucleases*. Cell, 1997. **91**(4): p. 457-466.
2. Mitchell, P., E. Petfalski, and D. Tollervey, *The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism*. Genes Dev, 1996. **10**(4): p. 502-513.
3. Zinder, J.C. and C.D. Lima, *Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors*. Gene & Development, 2017. **2**(31): p. 88-100.
4. Januszyk, K. and C.D. Lima, *The eukaryotic RNA exosome*. Curr Opin Struct Biol, 2014. **24**: p. 132-40.
5. Fasken, M.B., et al., *The RNA Exosome and Human Disease*. Methods Mol Biol, 2020. **2062**: p. 3-33.
6. Allmang, C., et al., *Functions of the exosome in rRNA, snoRNA and snRNA synthesis*. EMBO J, 1999. **18**(19): p. 5399-5410.
7. Schneider, C., et al., *Transcriptome-wide analysis of exosome targets*. Mol Cell, 2012. **48**(3): p. 422-433.
8. Pefanis, E., et al., *Noncoding RNA transcription targets AID to divergently transcribed loci in B cells*. Nature, 2014. **514**(7522): p. 389-393.
9. Houseley, J., J. LaCava, and D. Tollervey, *RNA-quality control by the exosome*. Nat Rev Mol Cell Biol, 2006. **7**(7): p. 529-39.
10. Klauer, A.A. and A. van Hoof, *Degradation of mRNAs that lack a stop codon: a decade of nonstop progress*. Wiley Interdiscip Rev RNA, 2012. **3**(5): p. 649-660.
11. Rigby, R.E. and J. Rehwinkel, *RNA degradation in antiviral immunity and autoimmunity*. Trends Immunol, 2015. **36**(3): p. 179-88.
12. Allmang, C., et al., *The yeast exosome and human PM-Scl are related complexes of 3'→5' exonucleases*. Genes Dev, 1999. **13**(16): p. 2148-58.
13. Weick, E.M., et al., *Helicase-Dependent RNA Decay Illuminated by a Cryo-EM Structure of a Human Nuclear RNA Exosome-MTR4 Complex*. Cell, 2018. **173**(7): p. 1663-1677.
14. Molleston, J.M., et al., *A conserved virus-induced cytoplasmic TRAMP-like complex recruits the exosome to target viral RNA for degradation*. Genes & Development, 2016. **30**(14): p. 1658-1670.
15. Schneider, C. and D. Tollervey, *Threading the barrel of the RNA exosome*. Trends Biochem Sci, 2013. **38**(10): p. 485-93.
16. Kögel, A., et al., *The human SKI complex regulates channeling of ribosome-bound RNA to the exosome via an intrinsic gatekeeping mechanism*. Mol Cell, 2022. **82**(4): p. 756-769.e8.
17. Wan, J., et al., *Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration*. Nat Genet, 2012. **44**(6): p. 704-708.
18. Morton, D.J., et al., *The RNA exosome and RNA exosome-linked disease*. RNA, 2018. **24**(2): p. 127-142.
19. Namavar, Y., et al., *Classification, diagnosis and potential mechanisms in pontocerebellar hypoplasia*. Orphanet J Rare Dis, 2011. **6**: p. 50.

20. Burns, D.T., et al., *Variants in EXOSC9 Disrupt the RNA Exosome and Result in Cerebellar Atrophy with Spinal Motor Neuronopathy*. *Am J Hum Genet*, 2018. **102**(5): p. 858-873.
21. Somashekar, P.H., et al., *Bi-allelic missense variant, p. Ser35Leu in EXOSC1 is associated with pontocerebellar hypoplasia*. *Clinical Genetics*, 2021. **99**(4): p. 594-600.
22. Carithers, L.J., et al., *A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project*. *Biopreserv Biobank*, 2015. **13**(5): p. 311-9.
23. Kim, D.E., D. Chivian, and D. Baker, *Protein structure prediction and analysis using the Robetta server*. *Nucleic acids research*, 2004. **32**(suppl_2): p. W526-W531.
24. Wegler, C., et al., *Global variability analysis of mRNA and protein concentrations across and within human tissues*. *NAR Genomics and Bioinformatics*, 2020. **2**(1): p. lqz010.
25. Q, T., et al., - *Integrated genomic and proteomic analyses of gene expression in Mammalian cells*. - *Mol Cell Proteomics*. 2004 Oct;3(10):960-9. doi: 10.1074/mcp.M400055-MCP200. Epub, (- 1535-9476 (Print)): p. T - ppublish.
26. Milligan, L., et al., *A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts*. *Mol Cell Biol*, 2008. **28**(17): p. 5446-57.
27. Falk, S., et al., *Mpp6 Incorporation in the Nuclear Exosome Contributes to RNA Channeling through the Mtr4 Helicase*. *Cell Rep*, 2017. **20**(10): p. 2279-2286.
28. Fraga de Andrade, I., C. Mehta, and E.H. Bresnick, *Post-transcriptional control of cellular differentiation by the RNA exosome complex*. *Nucleic Acids Research*, 2020. **48**(21): p. 11913-11928.

Figures

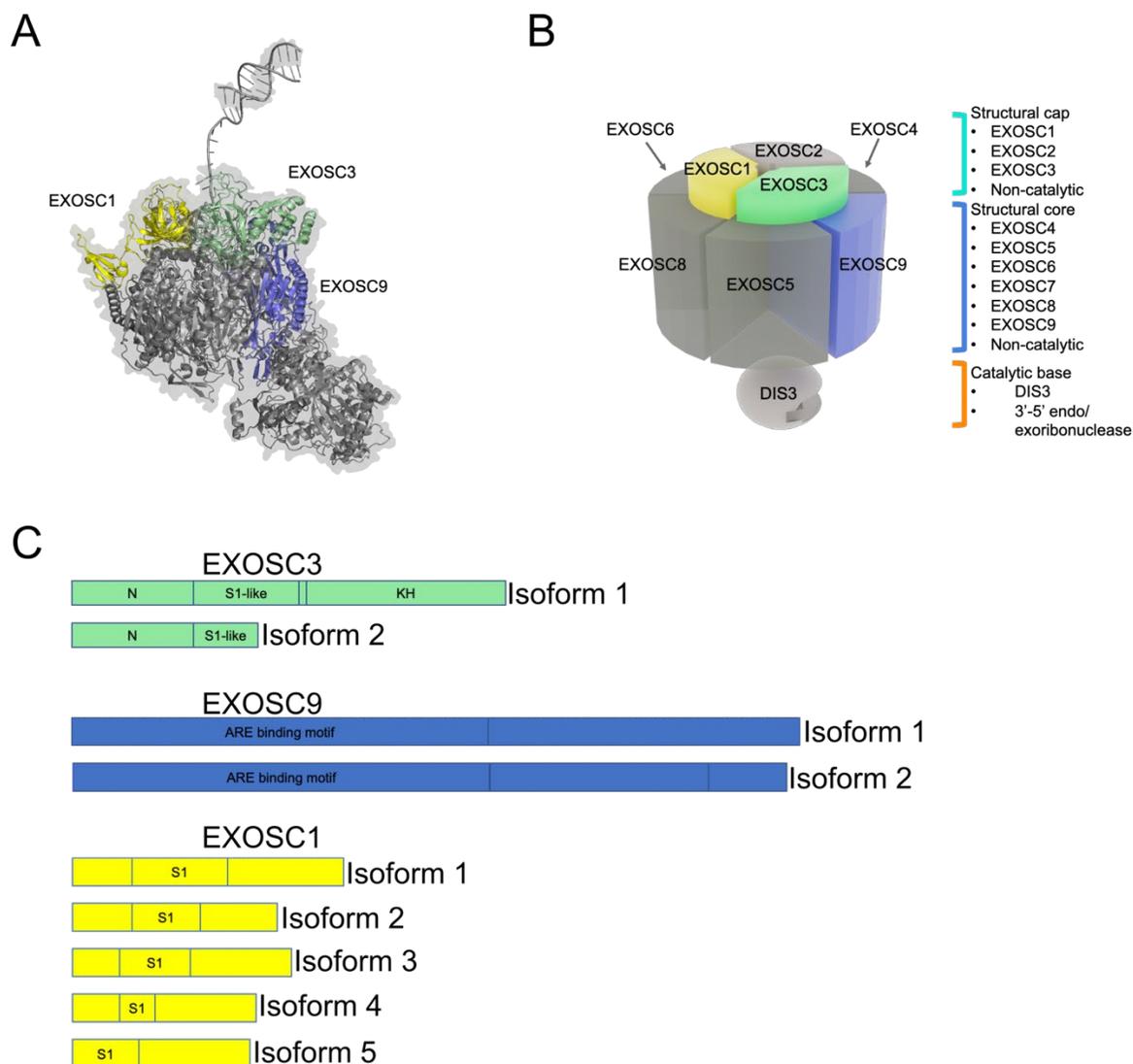


FIGURE 1. (A) A structure of the human RNA exosome complex (PDB #6D6Q) is shown [13]. (B) A cartoon rendering of the ten-subunit human RNA exosome complex inspired by PDB #6D6Q is presented including the three structural cap subunits (EXOSC1 (yellow), EXOSC2, and EXOSC3 (green)), six structural core subunits (EXOSC4, EXOSC5, EXOSC6, EXOSC7, EXOSC8, and EXOSC9 (blue)), and one catalytic subunit (DIS3). (C) Domain structures of human EXOSC3, EXOSC9, and EXOSC1 isoforms, including isoforms predicted by alternative splicing.

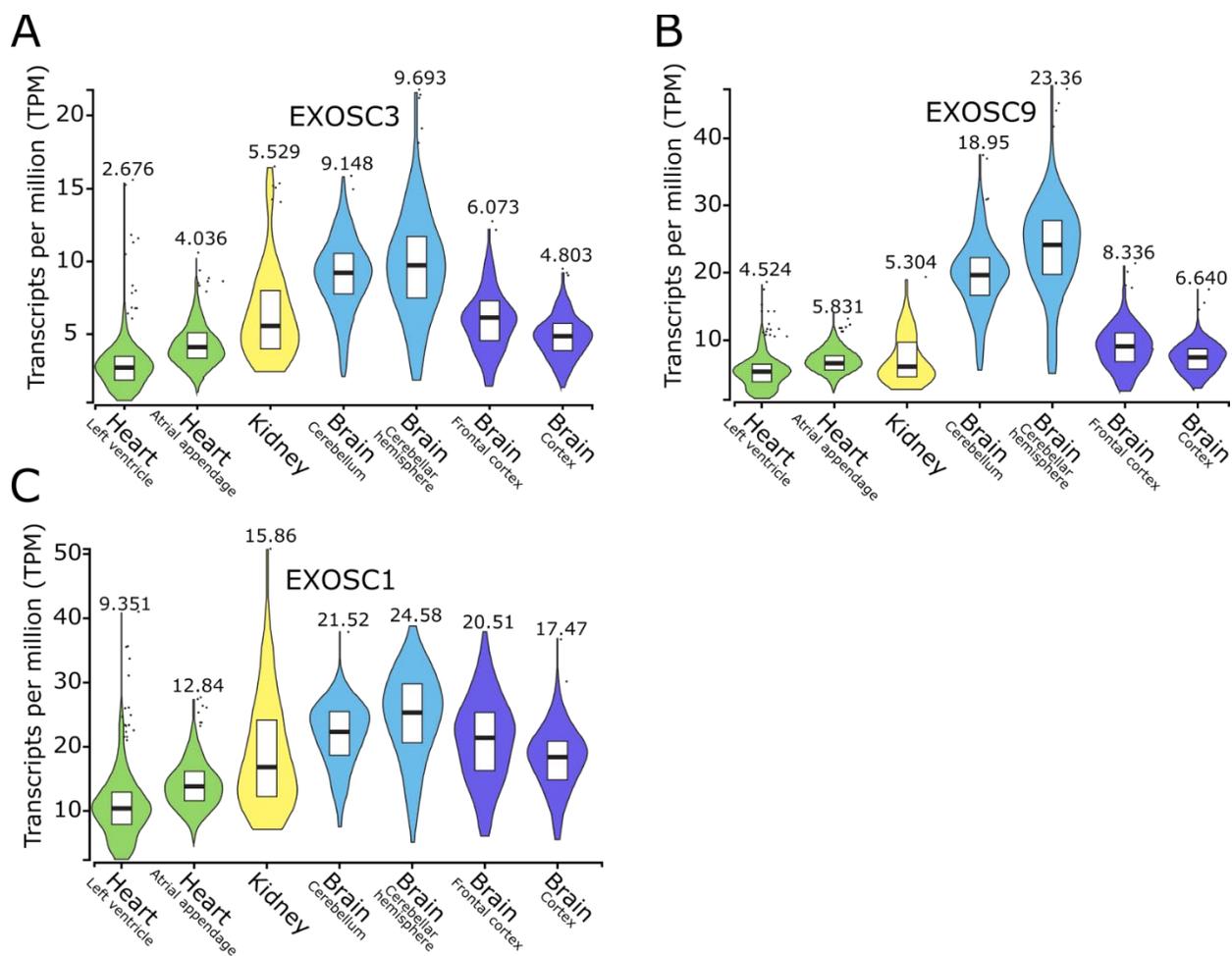


FIGURE 2. Steady state transcript levels for *EXOSC3* (A), *EXOSC9* (B), and *EXOSC1* (C) in heart (left ventricle and atrial appendage), kidney, and brain (cerebellum, cerebellar hemisphere, frontal cortex, and cortex). Median transcripts per million (TPM) values for *EXOSC3*, *EXOSC9*, or *EXOSC1* in respective tissues are indicated above each violin.

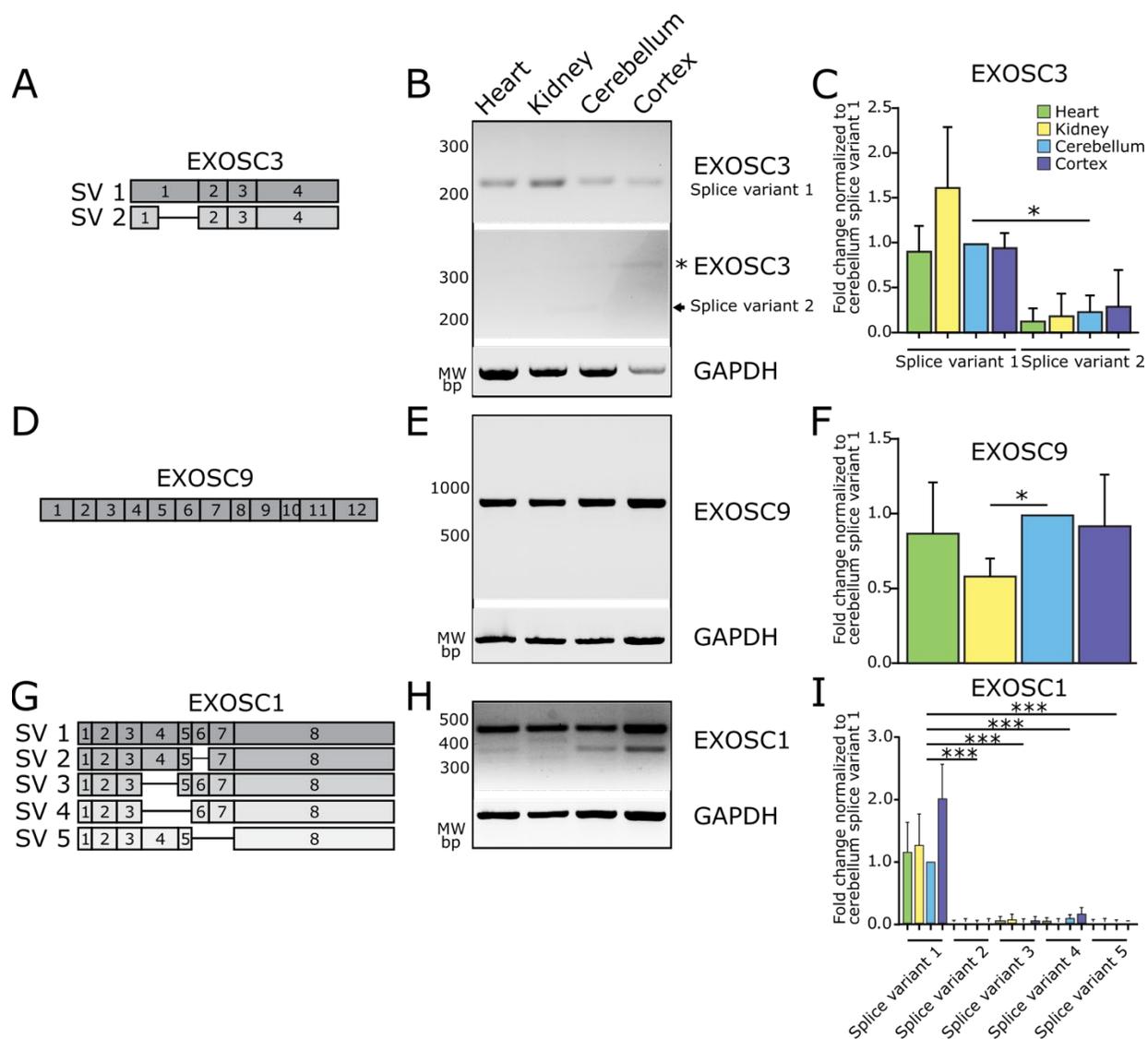


FIGURE 3. *Exosc3*, *Exosc9*, and *Exosc1* splice variants have different steady state levels in murine tissues. (A) Graphic depicting the two mouse splice variants (SV) of *Exosc3* available on NCBI. Each rectangle represents an exon. (B) RT-PCR of *Exosc3* splice variant 1 and splice variant 2 in murine heart, kidney, cerebellum, and cortex is resolved on an agarose gel. GAPDH is used as a loading control. The asterisk indicates a non-specific band. (C) Quantification of *Exosc3* RT-PCR shown in B. The fold change of *Exosc3* splice variant 2 in tissues is normalized to the fold change of *Exosc3* splice variant 1 in the cerebellum. (D) Graphic depicting the one murine splice variant (SV) of *Exosc9*

available on NCBI. Each rectangle represents an exon. (E) A single splice variant of *Exosc9* in murine heart, kidney, cerebellum, and cortex amplified by RT-PCR and resolved on an agarose gel. GAPDH is used as a loading control. (F) Quantification of *Exosc9* RT-PCR shown in E. The fold change is normalized by *Exosc9* present in the cerebellum. (G) Graphic depicting the five murine splice variants (SV) of *Exosc1* available on NCBI. Each rectangle represents an exon. (H) RT-PCR resolved on an agarose gel of *Exosc1* in murine heart, kidney, cerebellum, and cortex. GAPDH is used as a loading control. (I) Quantification of *Exosc1* RT-PCR shown in H. The fold change is normalized to *Exosc1* splice variant 1 present in the cerebellum. Error bars represent standard deviation. Statistical significance was calculated using one sample t-test. Significance is denoted by asterisk (* p-value < 0.05).

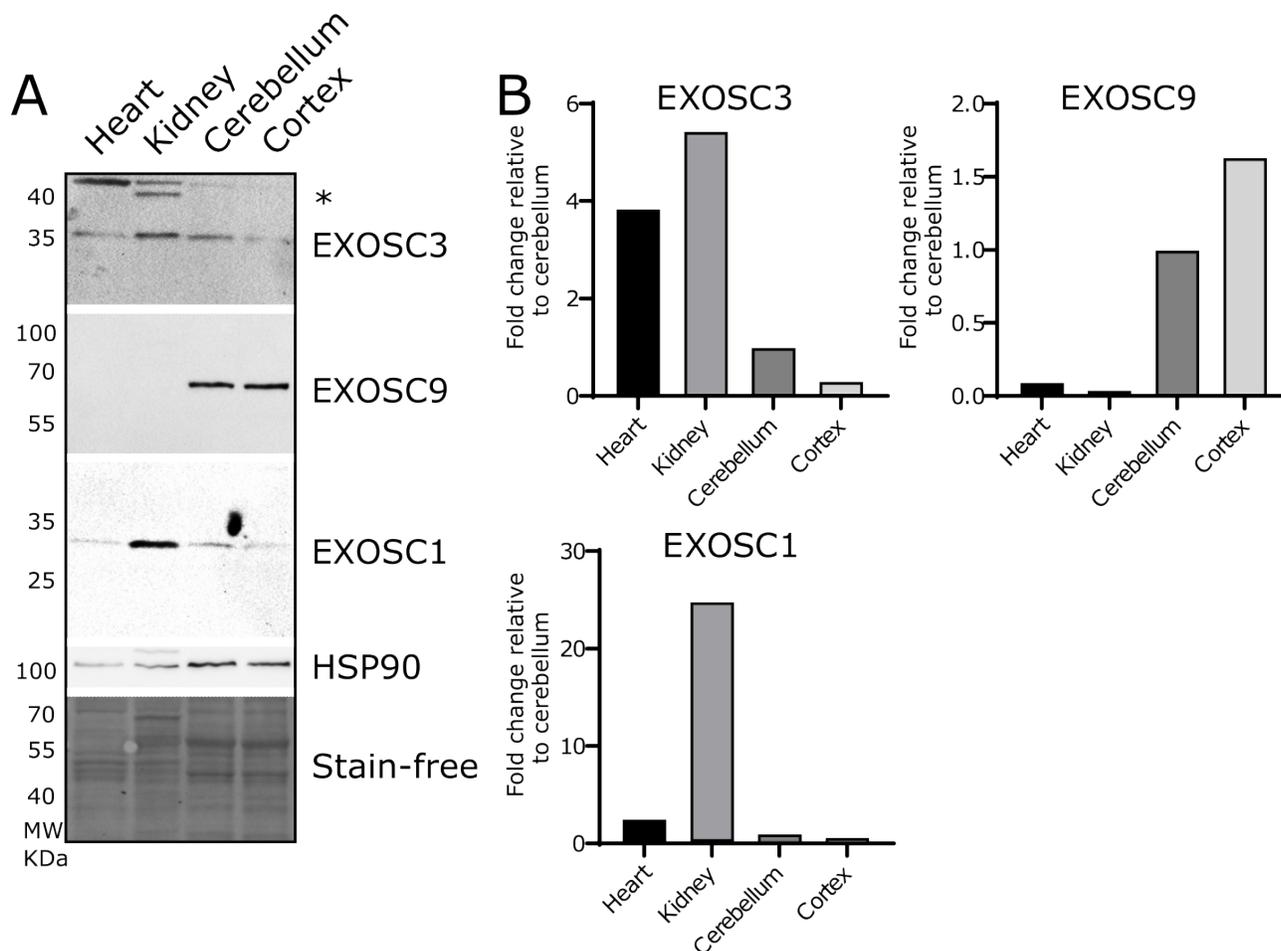


FIGURE 4. EXOSC3, EXOSC9, and EXOSC1 protein isoforms have different steady state levels in heart, kidney, cerebellum, and cortex. (A) Immunoblot of murine EXOSC3, EXOSC9, and EXOSC1. HSP90 and stain-free blots are used as loading controls. (B) Quantification of immunoblots shown in A. Fold change is calculated by normalizing to the protein levels of the cerebellum.

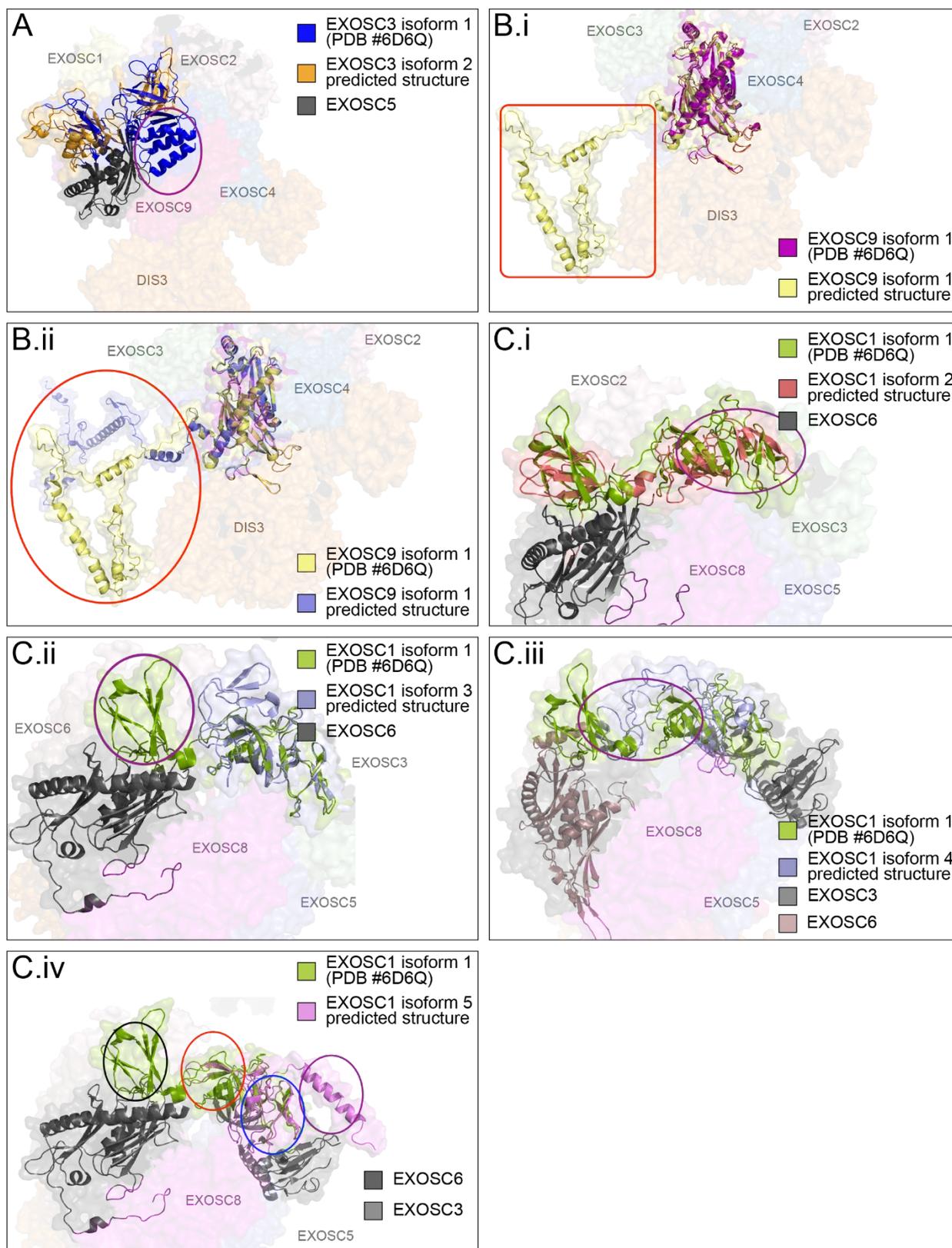


FIGURE 5. The human RNA exosome subunit isoforms differ in structure and interactions. Cryo-EM RNA exosome structure provided by Weick et al. (PDB #6D6Q) [13]. (A) The alignment between cryo-EM generated EXOSC3 isoform 1 (blue) and RoseTTAFold-generated EXOSC3 isoform 2 structure (gold). EXOSC3 isoform 2 lacks 16 amino acids (159-275) present in EXOSC3 isoform 1, which results in the absence of alpha helix cluster (circled in purple) of EXOSC3 isoform 1 in isoform 2. EXOSC5 interacts with EXOSC3 and is presented in grey. (B.i) The alignment between cryo-EM-generated EXOSC9 isoform 1 (purple) and the RoseTTAFold generated isoform 1 (yellow) structure. The cryo-EM-generated structure lacks a region (orange rectangle) present in the predicted structure. (B.ii) Alignment between RoseTTAFold-generated EXOSC9 isoform 1 structure (yellow) with RoseTTAFold-generated EXOSC9 isoform 2 structure (periwinkle). EXOSC9 isoform 2 lacks 17 amino acids (385-402) present in isoform 1. The differences between isoform 1 and isoform 2 are highlighted (orange circle). (C.i) Alignment between cryo-EM generated EXOSC1 isoform 1 (green) and RoseTTAFold generated isoform 2 (red). EXOSC1 isoform 2 lacks 42 amino acids (117-159) present in isoform 1, leading to structural differences (purple circle). Interacting subunit EXOSC6 is shown in grey. (C.ii) Alignment between cryo-EM-generated EXOSC1 isoform 1 (green) and RoseTTAFold-generated isoform 3 (grey-blue). EXOSC1 isoform 3 lacks 25 amino acids (50-75), and therefore does not contain the beta sheet structures (purple circle) that are present in isoform 1. Interacting subunit EXOSC6 is shown in grey. (C.iii) Alignment between cryo-EM-generated EXOSC1 isoform 1 (green) and RoseTTAFold-generated isoform 4 (light blue). EXOSC1 isoform 4 lacks 66 amino acids (49-115) that are present in the isoform 1 structure. The isoform 4 structure lacks beta sheets and misaligns with a

majority of the EXOSC1 structure represented by the purple circle. EXOSC3 and EXOSC6 are shown in light grey and grey-pink, respectively. (C.iv) Alignment between cryo-EM generated EXOSC1 isoform 1 (green) and RoseTTAFold-generated isoform 5 (pink). The isoform 5 predicted structure lacks 41 amino acids (118-159) that are present in the isoform 1 cryo-EM structure. The amino acids missing form beta sheets in isoform 1 represented by the black circle. Additionally, EXOSC1 isoform 5 forms an alpha helix structure (purple circle) that is not present in EXOSC1 isoform 1. Interacting subunits EXOSC6 and EXOSC3 are presented in grey and light grey.

Supplemental figures

RNA exosome subunit	Variant number	Accession number	Product size (bp)	Primer sequence
EXOSC3	Splice variant 1	NM_001362788.1	224	FWD: GCAGCAGAAGCGGTATGT REV: CCATCTCTGGTTCCATGTCTTT
	Splice variant 2	NM_025513.4	224	FWD: GGCCGTGCACAAGTATGTA REV: CCATCTCTGGTTCCATGTCTTT
EXOSC9	Splice variant 1	NM_019393.2	755	FWD: GTCCAAGGAGAGGAAGTAACAC REV: GCACTGGTCTGAGCTCTTATT
EXOSC1	Splice variant 1	NM_025644.4	481	FWD: GCAGCAGAAGCGGTATGT REV: CCATCTCTGGTTCCATGTCTTT
	Splice variant 2	NM_001320231.1	473	
	Splice variant 3	NM_001320232.1	384	
	Splice variant 4	NM_001164561.1	350	
	Splice variant 5	NM_001320233.1	337	

TABLE S1. RNA exosome subunits and associated accession numbers for *Exosc3*, *Exosc9*, and *Exosc1*.

RNA exosome subunit	Isoform number	Accession number	Number of amino acids
EXOSC3	Isoform 1	NP_057126.2	275
	Isoform 2	NP_001002269.1	164
EXOSC9	Isoform 1	NP_001029366.1	456
	Isoform 2	NP_005024.2	439
EXOSC1	Isoform 1	NP_057130.1	195
	Isoform 2	NP_001305292.1	152
	Isoform 3	NP_001305291.1	384
	Isoform 4	NP_001305291.1	129
	Isoform 5	NP_001305294.1	109

TABLE S2. Accession numbers for human protein sequences used to generate isoforms of EXOSC3, EXOSC9, EXOSC1 RNA exosome subunits through RoseTTAFold modeling. Amino acid lengths and assigned names are shown.