ECD functions as a novel RNA-binding protein to regulate mRNA splicing

Mohsin Raza[1], Asher Rajkumar Rajan[1], Achyuth Kalluchi[1], Irfana Saleem[1], Benjamin B. Kennedy[1], Kishor K. Bhakat[1,2], Hamid Band[1,2,3*], M. Jordan Rowley[1,2*], Vimla Band[1,2*]

[1]Department of Genetics, Cell Biology, and Anatomy, College of Medicine, [2]Fred and Pamela Buffett Cancer Center, [3]Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, Nebraska.

***Corresponding Authors:** Email: vband@unmc.edu; Phone: (402)-559-8565; Fax: (402)-559-7328; Email: jordan.rowley@unmc.edu; Phone: (402)-836-9794; Fax: (402)-559-7328. hband@unmc.edu; Phone: (402)-559-8572; Fax: (402)-559-4651.

Running Title: ECD binds to RNA and regulates RNA splicing.

Abstract:

The human ecdysoneless protein (ECD) plays an essential role in the regulation of cell cycle and cell survival. ECD has been implicated in RNA splicing through its association with the protein components of splicing complex. Here, using electrophoretic mobility shift assay and mutational analysis, we demonstrate that ECD directly binds to RNA through amino acids 135-148. Using enhanced CLIP-seq analyses, we identified a large repertoire of mRNAs bound to ECD. RNA-seq analyses revealed that ECD depletion in cells leads to widespread RNA splicing aberrations associated with alterations in gene expression. Significantly, we demonstrate that ECD mediates mRNA splicing by directly binding to RNA sequences located near splicing sites in context-dependent manner. Mechanistically, ECD associates with key proteins of the U5 snRNP complex in an RNA-independent manner and directly binds to U5 small nuclear RNA (U5 snRNA). Importantly, ECD-U5 snRNA interaction is crucial for maintaining the expression of U5-specific proteins. Furthermore, RNA binding defective mutant of ECD fails to rescue downregulated levels of U5 snRNP components or cell proliferation block induced by ECD knockout. Collectively, we provide compelling evidence that ECD regulates RNA splicing by directly associating with RNAs, and the RNA binding activity of ECD is essential for its function.

Introduction:

The interaction of cellular proteins with RNAs regulates key cellular processes, including transcription, RNA processing and translation, that are essential for cells to survive and replicate (1). RNA binding proteins (RBPs) are generally ubiquitously expressed and are among the most abundant proteins in a cell (2). Most RBPs interact with RNAs through RNA-binding domains comprised of specific amino acid sequences such as the RNA recognition motif (RRM), zinc finger (Znf) domains, heterogeneous nuclear RNP K-homology domain (KH), and others (3). However, an expanding group of noncanonical RBPs lacking traditional RNA-binding motifs has also emerged that includes transcription factors (TFs), epigenetic regulators, and cell cycle regulatory proteins, significantly broadening the functional diversity of the RBPs (4). Numerous examples of such noncanonical RNA-binding proteins have been characterized in the literature (5-7), where their additional RNA-binding capabilities contribute to more complex and diverse cellular functions, highlighting their multifaceted regulatory potential. In the present study, we show that ECD protein, despite lacking any traditional RNA-binding domain, directly binds to RNA and that its binding to RNA is essential for its role in RNA splicing to regulate cell proliferation and survival.

Previously, we identified ECD protein as a human papillomavirus HPV16 E6–binding protein (8). Subsequently, ECD has emerged as a key player in physiological processes (9-11) and various aspects of cancer development and progression (12-15). ECD overexpression in several types of cancers, including cervical and head & neck cancer (13), pancreatic cancer (16), gastric cancer (15,17), and breast cancer (18,19) is associated with poor prognosis and shorter patient survival. The co-oncogenic function of ECD was shown with mutant H-Ras for full transformation of mammary cells (12) and with HPV E7 to immortalize normal keratinocytes (13). Notably,

mouse mammary epithelium-targeted transgenic overexpression of human ECD led to the development of mammary hyperplasia and tumors, demonstrating a role of overexpressed ECD in tumorigenesis (14).

Physiologically, several studies using mutational, deletional, and knockdown approaches have demonstrated the vital roles of ECD in maintaining cell survival and cell cycle progression. Mutations or deletion of *Drosophila* Ecd leads to developmental and cell survival defects (20,21). Germline deletion of the mouse *Ecd* gene led to early embryonic lethality and conditional KO of *Ecd* in $Ecd^{flox/flox}$ mouse embryonic fibroblasts, led to G1 cell cycle block that was rescued by ectopic human ECD expression (9). ECD's role in cell survival was demonstrated by mitigation of ER stress by upregulating the levels of ER chaperone GRP78 (11).

More recently, ECD was found to associate with RNA biogenesis machinery proteins DDX39A (19) to regulate nuclear mRNA export and with PRPF8 (19,21) to regulate mRNA splicing. Notably, *Drosophila* Ecd was found to get incorporated into and promote the assembly of the U5 snRNP complex (22). While a crystal structure of ECD has not been solved, our circular dichroism and small-angle X-ray scattering analyses suggested a folded structure of the N-terminal residues (aa 1-432) suitable for a potential scaffolding function to organize protein-protein or protein–RNA interactions (23). Given the critical role of ECD in RNA biogenesis, we hypothesized ECD may directly bind to cellular RNAs to affect RNA biogenesis.

In this study, we discover and characterize ECD as a novel RNA-binding protein and identify the mechanism of its role in RNA splicing through direct binding to U5 snRNA and regulating the expression of U5 snRNP components. Importantly, using an RNA binding defective (RBD) mutant form of ECD, we demonstrate a requirement of RNA binding function of ECD in mRNA splicing, as well as in cell proliferation and survival.

Materials and Methods:

Cell lines and culture conditions

Immortalized mammary epithelial cell line, 76NTERT was cultured in DFCI-1 medium as described previously (24). Ecd$^{fl/fl}$ mouse embryonic fibroblasts (MEFs) were established previously (9), and maintained in Dulbecco's modified Eagle's medium (DMEM-Gibco#11965-09) supplemented with 10% fetal bovine serum (FBS-Gibco#10437-028), 10 mM HEPES (Hyclone#SH30237.01), 1 mM each of sodium pyruvate (Corning#25-000-CI), nonessential amino acids (Hyclone#SH30238.01), L-glutamine (Gibco#25030-081) and 50 µg/ml of gentamicin (Gibco#15750-078). The Lenti-X 293T cell line used for lentiviral packaging was obtained from Takara Bio (#632180) and grown in the complete DMEM medium described above. All cell lines were maintained in a humidified atmosphere with 5% CO_2 at 37°C. Routine mycoplasma testing was performed for all the cell lines using a mycoplasma detection kit (Sigma#MP0035).

Plasmid constructs and site-directed mutagenesis.

The human ECD ORF sequence was cloned into BamHI and XhoI sites of pGEX-6p-1vector for the expression of recombinant glutathione S-transferase (GST) fusion protein of full length ECD. The GST-fused truncated mutants of ECD (1-155, 150-438, and 439-644) have been described previously (9). pGEX-6p-1 constructs carrying internal deletions within ECD (Δ135-148, Δ263–276, Δ636–644, and Δ615–625) were custom-synthesized through Gene Universal, Newark, USA. For the expression of FLAG tagged ECD in mammalian cells, a lentiviral mammalian expression vector (pLV-Puro-EF1A) carrying FLAG tag at the N-terminus of wild

type (WT) human ECD was purchased from VectorBuilder Inc. Chicago, USA. A deletion mutant (Δ135-148) of ECD in the lentiviral expression vector was created using a PCR-based site-directed mutagenesis kit (NEB#E0554S), according to the manufacturer's instructions. The primer sequences used for mutagenesis are listed in supplementary Table S1. All constructs were verified by DNA sequencing.

Structure visualization of ECD.

PDB files of predicted 3D structures of human ECD (AF-O95905-F1-v4) and *Drosophila* Ecd (AF-Q9W032-F1-v4) proteins were obtained from the AlphaFold protein structure database (25). Visualization and superimposition of structures were performed using the PyMOL tool.

Generation of cell lines stably overexpressing WT or Δ135-148 mutant ECD.

Lenti-X 293T cells were co-transfected with lentiviral packaging plasmid, psPAX2 (Addgene#12260), envelope plasmid pMD2.G (Addgene#12259), and either empty lentiviral vector or vector carrying ECD (WT and mutant) using X-tremeGENE HP DNA transfection reagent (Roche#06366236001). 76NTERT cells were incubated with the lentiviral supernatants in the presence of polybrene (8 μg/ml, Sigma#H9268). 72 hours post-transduction, cells were selected in 1 μg/ml puromycin for 7 days, and ECD levels were assessed by Western blotting (WB) using an anti-ECD monoclonal antibody (9).

Generation of CRISPR-Cas9 mediated inducible ECD Knockout cell lines; Cre recombinase mediated KO in *Ecd*$^{fl/fl}$ MEFs.

To generate inducible ECD KO cell lines, lentiviral supernatants were prepared as described above using the Edit-R inducible lentiviral hEF1α-Blast-Cas9 plasmid (Horizon Discovery#CAS11229) and were used to transduce 76NTERT cells, followed by selection of Cas9-expressing cells in 5 μg/ml blasticidin for 7 days. These cells were then transduced with Edit-R lentiviral plasmids carrying synthetic guide RNAs against non-targeting control (NTC), (#GSG11811) or those targeting ECD: sgRNA1 (#GSGH11838-246505144; targeting sequence: TCCAAAGTCTCAACCCACAA) and sgRNA2 (#GSGH11838-246505139; targeting sequence: CTTGGGTATACTTACCCTAT). Following further selection in medium containing 5 μg/ml blasticidin and 1 μg/ml puromycin for 7 days, inducible loss of ECD in cell lines was assessed by WB after 96 hours of induction with 1μg/ml doxycycline (Sigma# D9891).

To conditionally KO Ecd in *Ecd*$^{fl/fl}$ MEFs, cells were infected with adenoviruses encoding mCherry-Cre or mCherry (control) procured from University of Iowa viral vector core. 96 hours post infection, cell lysates were collected, and WB was performed to confirm the KO of *ECD*.

Western blotting.

Cell lysates were prepared in RIPA buffer, supplemented with protease (Thermo Scientific#78429) and phosphatase (Thermo Scientific#78420) inhibitor cocktails and protein concentration was measured using the bicinchoninic acid (BCA) protein assay reagent (Thermo Scientific#23227). Western blotting was performed as described previously (10,19) using the following primary antibodies: anti-ECD mouse monoclonal antibody (9,18), anti-RB (Cell Signaling#9309S), anti-RUVBL1 (Cell Signaling#12300S), anti-DDX39A (Abcam#ab176348), anti-PRPF8 (Abcam#ab79237), anti-SNRNP200 (Bethyl#A303-454A), anti-EFTUD2 (Abcam#ab188327) and anti-β-actin (Sigma#A5441).

Expression and purification of GST tagged recombinant proteins.

The recombinant GST fusion proteins were expressed in bacteria and purified, as described previously (9,19). Briefly, the E. coli BL21 (DE3) transformed with the pGEX-6p-1 constructs of ECD were grown from single colonies in LB medium at 37°C to an OD600 of 0.6, followed by recombinant protein induction with 0.5 mM IPTG for 3 hours at 37°C. Bacterial lysates in sarkosyl-based lysis buffer were subjected to affinity chromatography on glutathione-Sepharose 4B beads (cytiva#17075601). The GST tag was cleaved using PreScission protease (cytiva#27084301) and removed with glutathione-Sepharose 4B beads.

***In vitro* GST fusion protein pull-down and co-immunoprecipitation.**

GST fusion protein pull-downs were performed, as described previously (10,19). Briefly, 1 mg protein aliquots of 76NTERT cell lysate in Triton-X100 based lysis buffer (50 mM Tris-HCl [pH 7.5], 150 mM NaCl, 0.5% Triton X-100) were incubated with 20 μg of glutathione Sepharaose-4B bead-bound WT GST-ECD, GST-ECD Δ135-148 mutant or GST alone for 4 hours at 4°C. Post incubation, samples were washed five times, and bound proteins were eluted by boiling the samples in SDS PAGE sample loading buffer for 10 minutes. The bound proteins were detected by WB using the indicated antibodies. Membranes were stained with Ponceau S (Sigma#P7170) to confirm the use of comparable amounts of GST fusion proteins. For co-immunoprecipitation (co-IP) experiments, cell lysates were prepared in Triton-X100 based lysis buffer supplemented with protease and phosphatase inhibitors. For each co-IP, 1mg of protein lysate was incubated overnight at 4°C with 2.5 μg of ECD (proteintech#10192-1-AP) or IgG antibody (Sigma#PP64B). Next, the protein-antibody complexes were captured by incubating the

samples with protein A Sepharose beads (Thermo Fisher#101141) for an additional 3 hours at 4°C. Post incubation, samples were washed 5 times in lysis buffer and western blotting was performed as mentioned above. For RNase treatment, protein lysate was treated with RNase A (Thermo Fisher #EN0531) at 10 µg/mg of protein for 15 minutes at room temperature (RT).

Electrophoretic mobility shift assay (EMSA) for ECD-RNA binding.

RNA EMSA was performed to assess the direct interaction of ECD with RNA using the LightShift Chemiluminescent RNA EMSA kit (Thermo Scientific#20158). Briefly, recombinantly purified WT ECD or its various mutants were incubated with 40 nM of biotinylated RNA probes (synthesized by IDT Technologies-Probes 1 and 2/ GenScript-Biotinylated U5 snRNA) for 45 minutes at RT in a total reaction mixture of 20 µL containing 1X RNA EMSA binding buffer (10 mM HEPES, 20 mM KCl, 1 mM MgCl2, 1 mM DTT) supplemented with 5% glycerol, 100 µg/mL tRNA (Thermo Scientific#20159) and 10 mM KCl. For antibody competition assay, ECD monoclonal antibody or control IgG was preincubated with purified ECD for 15 minutes in RNA EMSA buffer before the addition of biotinylated RNA probe. Following the incubation period, five microliters of 5× loading buffer was added to the 20 µL reaction mixture, and the samples were run on 7.5% TBE native polyacrylamide gel (pre-run at 100 V for 45 minutes in cold 0.5X TBE buffer) at 100 V until the sample dye migrated 3/4 of the length of the gel. Resolved samples were transferred to a positively charged nylon membrane (cytiva#NW0473) in 0.5X TBE at 60 V for 60 minutes. The membranes were crosslinked with UV light using the optimal-cross link function of the Spectroline UV crosslinker instrument. Signals were visualized using the detection reagents of the biotin-labeled RNA chemiluminescence kit, as instructed by the manufacturer.

RNA Immunoprecipitation.

Immunoprecipitation (IP) of ECD-associated RNAs was performed using the IP validation and RNA imaging kit (ECLIPSEBIO#143663), as per the vendor's instruction manual. Briefly, 76NTERT cells expressing vector or N terminal FLAG tagged ECD were UV-crosslinked at 254 nm wavelength (400 mJoules/cm^2) using a Spectroline UV crosslinker instrument. Protein lysates from crosslinked 76NTERT cells were subjected to IP using 5 µg each of a polyclonal anti-ECD (proteintech#10192-1-AP) or a monoclonal anti-FLAG (Sigma#F1804) antibody or their respective IgG-rabbit (Sigma#PP64B) and IgG-mouse (Sigma#CS200621) control antibodies. The transcripts pulled down by the indicated antibodies were end-repaired and ligated to a biotin adapter at the 3' end. The IP samples were run on NuPAGE 4-12% Bis-Tris protein gels, transferred to nitrocellulose membrane, and visualized using the chemiluminescent nucleic acid detection method. Concurrently, WB was performed to determine the pull-down efficiency of the antibodies. For WB, IP samples and 3% of the input samples were run on SDS-PAGE gels, transferred to PVDF membranes, and probed with the anti-ECD antibody.

eCLIP seq and data analysis.

The eCLIP analyses were performed under contract by Eclipse Bioinnovations Inc, San Diego, according to the published single-end seCLIP protocol (26) with the following modifications. 20 million cells were UV crosslinked at 254 nm wavelength (400 mJoules/cm^2) and then lysed using 1 mL of eCLIP lysis mix. Lysed samples were twice sonicated for 4 min with 30 seconds ON/OFF cycles at 75% amplitude (Q800R2 Sonicator, QSonica). 5 µg of polyclonal anti-ECD (proteintech#10192-1-AP) antibody pre-coupled to anti-Rabbit IgG Dynabeads was

incubated overnight with samples containing 50-100 μg of RNA at 4°C. The beads were washed with eCLIP high stringency wash buffers and collected using magnetic separation. The IP samples and 2% of paired input samples were resolved on gels, transferred to the membrane, and the region of the membrane from the ECD size to 75 kDa above was used to isolate the IPed RNA species. Subsequent RNA adapter ligation, reverse transcription, DNA adapter ligation and PCR amplification were performed as reported previously (26). Sequencing primers are ligated at 3'end and the eCLIP cDNA adapter containing a random sequence of 10 nucleotides at the 5' end, served as a unique molecular identifier (UMI) (27). UMIs were trimmed from read sequences using the umi_tools (v0.5.1) (28). The UMI sequences were added to the read names in the FASTQ files to be utilized for subsequent analysis. The 3'adapters were trimmed using cutadapt (v2.7) (29) and reads shorter than 18 bp were discarded. Next, reads were mapped to human repetitive elements and rRNA sequences from Dfam (30) and Genbank (31) databases. STAR (v2.6.0c) (32) was used to map all non-repeat reads to the human genome (hg38). Utilizing UMI sequences from the read names and mapping positions, PCR duplicates were removed using Umi tools (v0.5.1). Peaks were called using CLIPper (33). For each peak, IP versus input fold enrichment scores were computed as a ratio of counts of reads overlapping the peak region in the IP and the input samples. P-value was computed for each peak by the Yates' Chi-Square test or by the Fisher Exact test if the expected or observed read number was below 5. Different sample conditions were compared in the same manner as IP versus input enrichment. Next, for each peak (one sample type), enrichment and p-values relative to the normalized read counts overlapping these peaks in another sample type were calculated. Annotation of Peaks was done using the transcript information from GENCODE (34). To define the final annotation of the overlapping features, the priority hierarchy of CDS, UTRs, and Introns was considered, and the expected values were calculated by permutations or

random peaks. Motifs from eCLIP peaks were identified using HOMER (35) and the top 10 motifs were clustered and summarized using STAMP (36). ECD eCLIP peaks were compared to peaks identified for 149 other RBPs (37), liftOver was used to convert to the hg38 genome, and an enrichment score was calculated based on the ratio of genes targeted by both the annotated RBP and ECD vs. those targeted by ECD alone. Previously annotated functional categories were then summarized using the difference in average enrichment score for RBPs inside vs outside each category.

RNA-seq and mRNA Splicing Analysis.

Total RNA was isolated from control (sgNTC) or ECD depleted (sgECD-2, labeled as sgECD throughout the manuscript unless specified) 76NTERT cells using the Trizol reagent (Invitrogen#15596018) and the quality of purified RNA was determined using a Bioanalyzer. RNA-seq libraries were prepared using the TruSeq RNA Library Prep Kit v2 (Illumina) following the manufacturer's protocols and subjected to 100 bp paired-end sequencing on an Illumina NovaSeq 6000 system at the UNMC Next Generation Sequencing (NGS) facility. The NGS reads from the NovaSeq 6000 server were downloaded in FASTQ format and mapped to the human genome (hg38) using STAR (v2.7.3a) (32). Differentially expressed genes (DEGs) were identified using DESeq2 (v1.44.0) (38). The normalized Transcripts Per Million (TPM) values were derived using STRINGTIE (v2.1.1), (39) and differential splicing genes (DSGs) were assessed using rmats-turbo (v4.3.0) (40). Function enrichment/gene ontology (GO) over-representation analysis was performed using ENRICHR (41). Signal tracks were created using deeptools (v3.5.1), (42) with BPM normalization and visualization using IGV (43). Putative transcription factor (TF) target

genes were identified from the ARCHS4 database (44). TFs within the database were overlapped with DSGs from this study, and their targets were overlapped with DEGs.

Quantitative real-time PCR (qRT-PCR).

1µg of TRIzol-purified RNA was reverse transcribed using the iScript gDNA clear cDNA synthesis kit (Bio-Rad#1725035). qRT-PCR was conducted in a QuantStudio 3 real-time PCR system (applied biosystems) using the SYBR green master mix (Applied Biosystems#4309155). The list of primers used is provided in the Supplementary Table-S1.

CellTiter-Glo based luminescent cell proliferation assay.

Ecd$^{fl/fl}$ MEFs expressing the vector, WT ECD, or its Δ135-148 mutant were transiently infected with mCherry vector (control) or mCherry-Cre-expressing adenoviruses. 500 cells were plated per well in 96-well plates and cultured with a change of medium on alternate days. At the indicated time points, viable cells were quantified using the CellTiter-Glo 2.0 cell viability assay (Promega# G9241) according to the manufacturer's instructions.

Statistical analysis.

The results are expressed as mean ± standard error of mean (SEM) from three independent experiments, performed at least in triplicates each time. Statistical analyses were done using the GraphPad Prism 11 software. The type of statistical test used in individual experiments is indicated in figure legends. A p-value of ≤0.05 was considered statistically significant.

Results:

ECD directly binds to RNA through its N-terminal region

In previously published analyses of ECD-associated proteome, we identified multiple proteins involved in RNA biogenesis, including the mRNA splicing-associated, U5 small nuclear ribonuclear protein 200 (U5 SnRNP200), elongation factor Tu GTP binding domain containing 2 (EFTUD2), spliceosome scaffold protein PRPF8, and the mRNA export protein DDX39A (19). Additionally, *Drosophila* Ecd interaction with Prp8 (ortholog of mammalian PRPF8) was shown to regulate alternative mRNA splicing (21) through Prp8 stability by Ecd leading to the promotion of U5 snRNP maturation (22). Given the critical role of ECD in RNA biogenesis, we speculated that ECD may directly bind to RNA. The online tool RNABindRPlus (45) predicted three regions in ECD (aa 263-276, aa 615-625, and aa 636-644) with high scores for potential RNA binding **(Fig. 1A)**. Additionally, three other sites (aa 135–148, aa 205-219, and aa 301-309) in the N-terminal region were identified with lower prediction scores (**Fig. 1A).** We therefore tested the RNA binding potential of ECD by performing RNA EMSA using purified ECD protein and a biotinylated single-stranded mRNA probe, previously reported to bind DDX39A (46). The GST tag of the E. coli-purified recombinant ECD fusion protein was cleaved with precision protease and purity was confirmed by Coomassie staining of SDS-PAGE gels **(Fig. S1A)**. The RNA EMSA showed an ECD dose-dependent upward shift in the migration of the RNA probe signifying the formation of an ECD-RNA complex **(Fig. 1B)**. The specificity of binding was demonstrated by efficient competition by an excess amount of the unlabeled RNA probe **(Fig. 1B)**. Further, inclusion of an ECD specific monoclonal antibody eliminated the ECD RNA binding while the IgG control had no effect **(Fig. 1C)**. These observations supported the conclusion that ECD binds to RNA. To localize the region(s) needed for RNA binding, we conducted EMSA assays with

purified deletion fragments of ECD that included the N-terminal (aa 1-155), middle (aa 150-438), and C-terminal (aa 439-644) **(Fig. S1B)**. These analyses revealed that only the N-terminal region (aa 1-155) of ECD binds to RNA **(Fig. 1D)**. Interestingly, the RNABindRPlus predicted ECD N-terminal region (aa 135–148), though having a low score for RNA binding was the only predicted site within the aa 1-155 region **(Fig. 1A)**. Thus, to further validate our findings, we took an unbiased approach and purified ECD deletion mutants, either lacking aa 135-148 or other top three predicted amino acid regions (aa 263–276, 615–625 and 636–644) (**Fig. S1C)**. Notably, only ECDΔ135-148 mutant exhibited significant decrease in RNA binding **(Fig. 1E),** substantiating the conclusion that the N-terminal region of ECD binds to RNA and that aa 135-148 are required for this activity. Importantly, the AlphaFold-predicted structure of human ECD shows that aa 135-148 are part of anti-parallel beta sheets **(Fig. 1F),** and the corresponding region (aa163–173) in *Drosophila* Ecd adopts an identical secondary structure **(Fig. S2).** Consistent with a biologically key role of the RNA-binding region of ECD, the AlphaMisfold tool predicts high pathogenicity scores for amino acids around this region **(Fig. 1G).** Notably, anti-parallel beta-sheet structures are well-documented in the literature as potential RNA-binding domains (3). Together, our findings provide compelling evidence that ECD directly interacts with RNA through its N-terminal region and support the notion that RNA binding by ECD may underlie its biological roles.

ECD binds to cellular RNA

To assess the ECD's RNA binding ability in cells, we performed UV crosslinking and immunoprecipitation (CLIP) using 76NTERT cells stably expressing either vector or FLAG-tagged ECD. RNA species migrating over a large range on gels were detected in anti-FLAG IPs of FLAG-ECD-expressing but not in vector cells, and such species were detected in with or without

FLAG-tagged expressing cell lines with anti-ECD antibody, with more robust RNA-IP in FLAG-ECD expressing cells (**Fig. 2A**). As expected, control mouse or rabbit IgG antibody IPs did not show any co-IPed RNAs **(Fig. 2A).** Concurrent anti-FLAG **(Fig. 2B)** or anti-ECD **(Fig. 2C)** immunoblotting demonstrated the efficiency of IPs and lack of pull-down with respective control IgGs further assures the specificity. These results clearly support the notion that ECD binds to RNA within the cell.

Next, to directly assess cellular RNAs associated with ECD under physiological conditions, we performed enhanced UV crosslinking and IP followed by sequencing (eCLIP-seq) analysis (47) using an anti-ECD polyclonal antibody to IP ECD-associated RNAs in 76NTERT cells. The eCLIP-seq analysis revealed a total of 3,363 transcriptome-wide ECD binding sites that were shared between replicates based on irreproducible discovery rate (IDR) analysis using the CLIPper pipeline (33). A vast majority of ECD binding sites were primarily localized in coding exons (CDSs-71%), and 5′ untranslated regions (5'UTRs-20%) with minimal overlap with 3′UTRs (5%), and introns (4%) **(Fig. 2D)**. Monte Carlo permutation testing to control for relative feature sizes showed enrichment of ECD in both CDSs and 5'UTRs **(Fig. 2E).** De novo analysis of peak-associated motifs identified several significant sequences, summarized as GA(UC) rich sequences **(Fig. 2F)**. Notably, only 35% of ECD peaks overlapped these motifs, indicating that while ECD may prefer GA-rich sequences, binding to RNA is not associated with one specific sequence **(Fig. 2F)**. An EMSA experiment confirmed the ability of purified ECD to bind to the eCLIP-identified RNA motif, used as a biotinylated RNA probe, with efficient competition by excess unlabeled RNA probe **(Fig. 2G).** These results corroborate the eCLIP-seq findings, further solidifying that ECD functions as an RNA-binding protein.

GO biological process pathway analysis of ECD binding RNAs that possess the identified ECD-binding consensus motif showed enrichment for key cellular pathways including the transcription, apoptosis, response to endoplasmic reticulum stress and cell proliferation **(Fig. 2H).** These pathways play a crucial role in physiological functions and commonly altered in oncogenesis and tumor progression, consistent with previously identified roles of ECD in physiological context (9,11) and tumor initiation and progression (14,16).

ECD deficiency leads to transcriptome-wide alteration in gene expression and causes aberrant retention of spliced-out RNAs.

Given the broad repertoire of cellular mRNAs that we identified to be in complex with ECD, we performed RNA-seq analysis of control vs. ECD depleted 76NTERT cells using doxycycline-regulated Cas9 for short-term ECD loss since stable KO is not compatible with cell viability (9). Before RNA-seq analysis, loss of ECD was checked concurrently in cell lysates from doxycycline treated control (sgNTC) and sgECD groups. Data revealed a marked reduction in ECD protein levels upon doxycycline-induced depletion **(Fig. 3A).** RNA-seq analysis revealed that the replicates of both the control and ECD-depleted conditions were similar within their respective groups **(Fig. S3A)** and ECD depleted group showed distinct transcriptional patterns from control cells **(Fig. 3B).** Differential expression testing using log2 FC $\geq$ 1 and adjusted P-value < 0.05 as a cutoff, showed differential expression of 3,452 genes, with 1,317 downregulated and 2,135 upregulated genes in ECD depleted cells **(Fig. 3C).** Biotype analysis revealed that upon ECD depletion, DEGs corresponded to protein coding (34.4%), lncRNA (29.2%) and Pseudogene (26.4%) categories **(Fig. S3B).** MSigDB gene set analysis of DEGs showed that those downregulated upon ECD depletion are enriched in cholesterol homeostasis, E2F target, KRAS

signaling, and G2-M checkpoint pathways, whereas the upregulated genes are enriched in inflammatory response, TNF-alpha and hypoxia signaling pathways **(Fig. 3D).** GO biological process analysis also indicated an enrichment of downregulated genes in sterol and cholesterol biosynthetic processes along with protein-DNA and nucleosome assembly, while upregulated genes corresponded to regulation of type II interferon production **(Fig. 3E).** Heatmaps illustrate the downregulation of cholesterol homeostasis and protein-DNA complex assembly related genes **(Fig. 3F)** and upregulation of inflammatory response and negative regulation of type II interferon pathway genes **(Fig. 3G)** upon ECD depletion**.** These results support a key role of ECD in ensuring the expression of multiple gene networks involved in diverse physiological functions.

Given our (13) and others' (21,22) published studies linking ECD to regulation of mRNA splicing, we evaluated the impact of ECD depletion on the transcriptome-wide mRNA splicing. Based on an FDR of < .05 cutoff, we observed differential splicing pattern in ECD depleted cells **(Fig. 4A).** The rMATs tool identified a total of 5,070 DSEs in ECD depleted cells as compared to control cells and these splicing events map to 2,595 unique genes, suggesting many genes had multiple differential splice events **(Fig. 4B)**. Indeed, 1,844 genes showed one, 527 genes showed two, 141 genes showed three, and 83 genes showed four or more DSEs **(Fig. 4B).** Notably, exon skipping and intron retention were the most altered splicing events upon ECD depletion, while the impact on other splicing events like alternative 5' splice site and 3'splice site usage was less **(Fig. 4C).** ECD depleted cells showed downregulation at 1,373 exon skipping events out of the 5,070 total DSEs. In contrast, aberrant splicing involving intron retention was amplified, with 2,220 intron retention events out of 5,070 total identified DSEs **(Fig. 4C).** Interestingly, MSigDB pathway analysis showed an enrichment for mitotic spindle and DNA repair pathways among genes exhibiting altered splicing upon ECD depletion **(Fig. 4D).** GO molecular function analysis

revealed enrichment for RNA binding and cadherin binding functions for genes with altered splicing upon ECD depletion **(Fig. 4E).** Furthermore, GO biological process revealed enrichment for transcription and chromatin modeling pathways among the DSGs observed upon ECD depletion **(Fig. S4A).** RNA-seq signal tracks of representative genes belonging to top altered categories, *ARAP3* from mitotic spindle **(Fig. 4F)**, *MOV10* from RNA binding **(Fig. 4F),** and *CCNT2* from positive regulation of DNA templated transcription categories **(Fig. S4B)** illustrating aberrant RNA retention after ECD depletion. Overall, these results support the importance of ECD in regulating transcriptome-wide mRNA splicing events and transcript levels in mammalian cells to help maintain diverse and functionally critical molecular pathways.

ECD binding to RNAs mediate alternative splicing

To delineate if the RNA-binding activity of ECD is linked to its role in mRNA splicing and/or gene expression, we first analyzed the overlap between DEGs or DSGs upon ECD depletion. We found minimal overlap between 3,452 DEGs and 2,595 DSGs, with only 179 genes shared between the two groups **(Fig. 5A).** To probe this relationship more directly, we assessed the overlap of ECD eCLIP peaks with DEGs and DSGs vs. randomly selected negative control genes. While only 5.56% of DEGs overlapped with ECD eCLIP peaks, a significantly higher proportion, 22.19%, of DSGs showed such overlap **(Fig. 5B)**. Additionally, mapping the average profile of ECD eCLIP enrichment signals (Log2 IP/input) at differentially spliced coordinates revealed enhanced enrichment signals precisely at mis-spliced junctions **(Fig. 5C)**. Visualization of representative gene tracks, *AGRN* **(Fig. 5D)** and *MOV10* **(Fig. 5E)**, demonstrate ECD occupancy of the transcript that is mis-spliced after ECD depletion. These findings strongly

support the idea that ECD interacts with RNAs near their splicing sites to regulate alternative splicing events.

The weaker linkage of ECD's RNA binding activity with DEGs **(Fig. 5B)** suggested the likelihood that regulation of gene expression by ECD may not be through direct association with RNAs. We therefore speculated that ECD regulates gene expression indirectly by altering the splicing of genes that may encode transcriptional factors (TFs), affecting their target genes. Consistent with this speculation, out of 2,595 identified DSGs, 180 genes were found to encode TFs, and 32 of these were bound to ECD in our eCLIP-seq analyses. Concurrently, we analyzed the ARCHS4 database and identified 1,996 genes exhibiting differential expression upon ECD loss as direct targets of TFs in this database. Among these, 449 genes were regulated by TFs whose RNAs were directly bound to ECD and aberrantly spliced. **(Fig. 5F).** Ranking of TFs whose RNAs were bound to ECD and differentially spliced upon ECD depletion based on the number of their targets exhibiting differential expression (compared to randomly selected genes) identified four TFs, ELF3, EPAS1, BHLHE41, and IRF1 that showed significant enrichment of DSEs among their targets **(Fig. 5G).** These results suggest that ECD binds and splices RNAs of TFs to regulate the expression of diverse genes.

Next, we examined how the RNA repertoire bound by ECD compares with that of other well-established RNA-binding proteins to develop a comprehensive understanding of sequence specificity of RNAs bound to ECD and to do so we compared ECD eCLIP signal to that of 149 various RBPs from a reference eCLIP dataset (37). Interestingly, based on overlap enrichment ratio, the top 10 RBPs with RNA-binding profiles similar to ECD included PRPF8 and EFTUD2 **(Fig. 6A),** two U5 snRNP complex proteins which we found previously to interact with ECD (19). We then examined previously annotated functional categories that were assigned to RBPs and,

notably, “spliceosome” and “splicing regulation” emerged as the top-hits for RBPs that have similar RNA binding profiles to that of ECD **(Fig. 6B)**. Moreover, the average eCLIP signal plots for each of the top 10 spliceosome-associated RBPs around ECD dependent differential splicing events demonstrated that these proteins bind to RNA near sites that become mis-spliced in the ECD depleted cells **(Fig. 6C)**. These findings reinforce the potential involvement of ECD bound transcriptome in RNA processing and splicing-related mechanisms. Additionally, as an example of the overlap, we compared the eCLIP signal tracks for each of the top 10 RBPs within the spliceosome category at *TMEM214*, a representative gene that exhibits both ECD eCLIP signal and increased intron retention in ECD depleted cells. Our analysis revealed that ECD, like other RBPs annotated in the spliceosome category, displays a diffuse binding signal across the gene rather than a sharply defined single binding site **(Fig. 6D)**. This observation argues against a strictly sequence-driven binding model for ECD. Overall, these findings suggest a potential role for ECD in modulating RNA splicing through a broader, context-dependent mechanism.

ECD directly interacts with U5 snRNA and maintains the expression of key components of the U5 snRNP complex.

Our results that RNAs of only ~22% of DSGs in ECD depleted cells are directly bound to ECD, suggested additional mechanisms by which ECD regulates RNA splicing. Since we have shown that ECD associates with proteins that are components of the U5 snRNP complex (19) and others demonstrated that *Drosophila* Ecd regulates the stability of the U5 component Prp8 and also co-IPs with U5 snRNA (22), we assessed if the RNA binding activity of ECD may be linked to the regulation of U5 snRNP. First, we examined the impact of ECD depletion in cells on the levels of key U5 snRNP proteins. As reported, ECD depletion not only led to the downregulation

of PRPF8 levels (22) but also led to a marked reduction in the levels of other U5 snRNP proteins SNRNP200 and EFTUD2 **(Fig. 7A)**. Notably, our RNA sequencing analysis revealed reduced levels of *RNU5A-1*, gene that encodes U5 snRNA, the essential RNA component of the U5 snRNP complex, upon ECD depletion **(Fig. 7B)**. This finding was further validated through qPCR analysis of independent samples **(Fig. 7C).** We, therefore, investigated the possibility that ECD being an RNA binding protein, may directly interact with U5 snRNA. EMSA with a biotinylated full-length U5 snRNA as the probe demonstrated strong binding of WT ECD to U5 snRNA, whereas the RBD mutant exhibited markedly reduced binding. Moreover, the addition of a monoclonal anti-ECD antibody effectively abrogated the ECD-U5 snRNA interaction **(Fig. 7D).** GST pull-down assay demonstrated that the mutant retained its ability to bind to PRPF8, EFTUD2, DDX39A, RUVBL1 and RB proteins that we previously identified as ECD interacting proteins comparably to the WT ECD **(Fig. 7E),** ruling out any gross structural misfolding of the ECD protein upon deletion of amino acids 135–148. This result also indicates that ECD can associate with protein components of the U5 snRNP complex independently of RNA. To further validate this observation *in vivo*, we performed pulldown assays in 76NTERT cells with or without RNase A treatment. Notably, we observed comparable levels of U5-specific proteins pulled down by ECD in the presence or absence of RNA **(Fig. S5)**, confirming that ECD directly associates with the U5 snRNP complex, independent of RNA.

To further explore the biological significance of ECD's interaction with U5 snRNA, we utilized mouse embryonic fibroblasts (MEFs) harboring floxed Ecd alleles ($Ecd^{fl/fl}$), where the endogenous *Ecd* can be deleted by adenovirus-mediated Cre recombinase (9). As expected, removal of endogenous mouse *Ecd* with Cre adenovirus resulted in the downregulation of PRPF8, SNRNP200 and EFTUD2 protein levels **(Fig. 7F)**. Significantly, while Full length human ECD

restored the levels of PRPF8, SNRNP200 and EFTUD2 proteins, the RBD mutant was defective **(Fig. 7F)**, despite of its ability to pull down the U5 snRNP complex protein components from cell lysates (**Fig. 7E).** These findings strongly support the critical role of the ECD-U5 snRNA interaction in maintaining the expression of the U5 snRNP components.

The RNA binding activity of ECD is critical for its requirement for cell proliferation

We have shown that ECD is required for cell proliferation in diverse cellular contexts (9,11). To assess if the RNA binding activity of ECD is required for this function, we assessed the ability of WT or RBD mutant of human ECD to rescue *Ecd*$^{fl/fl}$ MEFs from proliferative block imposed by the deletion of endogenous mouse *Ecd* (9). KO of endogenous mouse *Ecd* and expression of exogenous human ECD or its mutant protein were confirmed through WB **(Fig. 8A)**. Although the level of the mutant ECD was lower than that of the WT ECD, it was about 1.4 times higher than the endogenous mouse Ecd **(Fig. 8A).** As expected, endogenous *Ecd* deletion in *Ecd*$^{fl/fl}$ MEFs resulted in a proliferation arrest, determined through measuring proliferation rate using cell-titre glo cell viability assay (**Fig. 8B)** and colony formation assay **(Fig. 8C & D).** Notably, this proliferation block was significantly rescued by WT human ECD expression (**Fig. 8B-D**). In contrast, ectopic expression of the RBD mutant (Δ135-148) failed to rescue the proliferation arrest **(Fig. 8B-D)**. These results support the conclusion that the RNA binding activity of ECD is required for its broader biological function in a cellular proliferation context.

Discussion

ECD is an evolutionarily conserved protein with critical developmental and physiological roles and whose overexpression contributes to promoting oncogenic behaviors in multiple cancers

(12-16,18). Recent studies have linked ECD to fundamental processes in RNA biogenesis, including the regulation of mRNA splicing through its interaction with, and regulation of the stability of, the U5 snRNP component PRPF8 (22), but the intrinsic biochemical activities of ECD that underlie its roles in mRNA splicing and the linkage of such activities to its cellular roles remains unclear. Here, we identify ECD as a novel RNA binding protein that directly binds to a large subset of cellular mRNAs as well as to U5 snRNA to regulate mRNA splicing. We demonstrate that ECD's RNA binding activity requires the structurally conserved N-terminal region sequences, and RBD mutant lacks the ability to stabilize the U5 snRNP complex or to rescue the cell proliferation. Further, we provide evidence that association of ECD to protein components of U5 snRNP complex is RNA independent. Our findings provide novel mechanistic insights into an essential cellular regulator of mRNA splicing whose biochemical roles have been nebulous thus far.

RNA splicing is critical to cellular function (48) and in addition a crucial ingredient in expanding and diversifying the translatome through alternative splicing (AS) (49). These mechanisms are central to many physiological processes, including development, differentiation and stress responses, and their dysregulation has been implicated in diseases like cancer, neurodegenerative disorders, and metabolic syndromes (50). RBPs play a pivotal role in regulating mRNA splicing as well as other key processes that control gene expression by directly interacting with target RNAs (51,52). Thus, our findings that ECD functions as an RNA binding protein and that this activity is linked to its cellular function provide novel mechanistic insights into how ECD functions in regulating RNA splicing.

Prior work demonstrating ECD's interaction with proteins involved in RNA export and splicing and its critical role in these functional processes (13,19) suggested the possibility that

ECD may directly interact with RNAs like other RNA biogenesis regulatory proteins (52). We provide multiple lines of evidence that this indeed is the case. These included in silico RNABindRplus prediction of potential RNA binding motifs on ECD, the N-terminal based RNA binding validated by RNA EMSA analysis **(Fig. 1A-D)**. Selective deletion of amino acids 135-148 in ECD significantly retarded its RNA binding activity **(Fig. 1E)** but did not alter its interactions with proteins known to be associated with other regions of ECD, including PRPF8, EFTUD2, DDX39A, RUVBL1 and RB **(Fig. 7E).**

While ECD's structure has not been determined experimentally, our circular dichroism (CD) and small-angle X-ray scattering (SAXS) studies (23) and Alpha-fold modeling (25) support a well-folded structure of the N-terminal region of ECD where the RNA binding critical amino acids are located **(Fig. 1F)**, whereas the C-terminus is disordered, containing intrinsically disordered regions (IDRs). Notably, IDRs are a characteristic of many RBPs where they confer structural flexibility to enhance the RNA interaction potential (53). RNA chaperones often utilize IDRs as functional domains for dynamic RNA interactions and to facilitate RNA folding or processing (54,55). While our analyses show that IDR containing C-terminal regions of ECD are dispensable for RNA binding, it remains possible that the IDR containing C-terminal region may facilitate RNA binding *in vivo*, potentially by enhancing ECD's conformational flexibility or interaction dynamics within the cellular environment or through protein-protein interactions given the previously defined interactions of ECD with proteins like PIH1D1 component of the R2TP co-chaperone complex and with the cell cycle regulator RB are mediated by C-terminal regions (9,10). Notably, while the *Drosophila* Ecd sequence (aa 160–173) corresponding to human ECD aa 135–148, which we identify as essential for RNA binding, do not exhibit high conservation at individual amino acids but both are predicted by the AlphaFold to form identical anti-parallel beta sheet

structures **(Fig. S2).** Given the known importance of anti-parallel beta sheets in RNA recognition motifs (RRM) of RNA binding proteins (56), it is likely that the region we identify as critical for RNA binding activity of human ECD will encode a conserved RNA binding activity across species.

Our eCLIP-seq analyses provided evidence for ECD's broad role as an RBP, demonstrating 3,363 unique transcriptome-wide binding sites on mRNAs encoded by genes involved in essential cellular processes, including transcription, apoptosis, endoplasmic reticulum stress, cellular differentiation, and proliferation **(Fig. 2H).** These results are consistent with key roles of ECD in physiological processes such as cell proliferation (9) and cell survival under endoplasmic reticulum stress (11) as well as the role of ECD overexpression in promoting tumorigenesis (13-16,19).

Notably, mapping the RNA binding preferences of ECD revealed a strong preference for coding sequences (CDS) and 5' untranslated regions (5'UTRs) of target RNAs. The top ECD binding motif present on a fraction of target RNAs identified as a purine rich sequence, GAGGAGGAGGAG, that includes the sequence of GAGGA, previously reported to present in exonic splicing enhancers (ESE) (57). ESE sequences are typically purine-rich and often consist of motifs alternating between adenosine (A) and guanosine (G), spanning six or more nucleotides. Their purine-rich composition is known to enhance the binding of splicing factors and facilitate RNA splicing (58). Consistent with the functional importance of ECD binding to splicing relevant motifs in target RNAs, depletion of ECD resulted in thousands of differential splicing events, especially causing alteration in the exon skipping and intron retention features **(Fig. 4A-C).** Significantly, exonic ESE sequences within viral and mammalian genes are known to play a critical role in regulating alternative RNA splicing (59). Notably, we previously found ECD to be

important in regulating viral and cellular RNA splicing in the Human papillomavirus (HPV) positive, cervical cancer cell line model (13).

A key finding of our study is the significant enrichment of ECD binding peaks in genes undergoing ECD-modulated splicing. We provide evidence that ECD binding sites are particularly enriched near mis-spliced junctions **(Fig. 5C),** strongly suggesting that ECD's RNA-binding function plays a direct role in regulating alternative splicing. However, only about 22% of the differentially spliced RNAs are directly bound to ECD, suggesting that ECD regulates splicing through additional mechanisms beyond direct binding to splicing sites or regulatory elements. One of these mechanisms is likely to be the requirement of ECD for proper splicing of transcription factors and accessory proteins that regulate transcription, splicing, stability, and other aspects of RNA biogenesis. In this regard, a significant subset of genes whose transcripts showed aberrant splicing were transcription factors whose target gene mRNA levels were altered (**Fig. 5F & G**). This explains one of the mechanisms through which ECD regulates the expression of thousands of genes, including both coding and non-coding RNAs **(Fig. 3C & S3B).** However, it also impacts gene expression for some of its targets through direct RNA binding, as transcripts of 6% of total DSEs were identified to bind directly to ECD. Our observation that the ECD-bound transcriptomic profile overlaps with the eCLIP-seq profiles of PRPF8 and EFTUD2 (**Fig. 6A-D**), two protein components of the U5 snRNP complex that associate with ECD suggests that the RNA binding specificity of ECD *in vivo* is not strictly RNA sequence-driven and potentially context dependent, influenced by protein interactome. The absence of strict RNA sequence specificity of ECD aligns with certain previously reported RBPs involved in splicing regulation, which exhibit minimal or no sequence selectivity, recognizing RNA structures or being influenced by factors such as protein cofactors (37,60).

We and others have previously reported that ECD engages in protein-protein interactions with components of the RNA biogenesis and translation machinery, with several components of the U5 snRNP in the ECD-associated proteome (19,22). Such protein-protein interactions have suggested one plausible mechanism for ECD to play a role in RNA splicing, with recent evidence that *Drosophila* Ecd functions as a chaperone to stabilize the U5 snRNP during its cytoplasmic assembly prior to nuclear import (22). We demonstrate that ECD directly interacts with U5 snRNA **(Fig. 7D),** the essential RNA component of the U5 snRNP complex (61), while its mutant with deletion of the critical residues involved in RNA binding fails to stabilize the U5 snRNP. Our results show that, like the drosophila Ecd (22), the mammalian ECD is required to help maintain the expression of the U5 snRNP complex protein component PRPF8, and we extend this to other U5 snRNP specific proteins including the SNRNP200, an RNA helicase and EFTUD2, a GTPase. Both proteins help dissociate the U4 and U6 snRNPs during splicing to facilitate the processing of precursor mRNAs into mature mRNAs (62,63). Importantly, while the WT ECD rescued the protein expression of PRPF8, SNRNP200 and EFTUD2 observed upon ECD loss, the RNA binding mutant ECD failed to do so **(Fig. 7F)**, thus establishing the RNA binding activity of ECD as pivotal to U5 snRNP biogenesis. It is notable that the prior work with *Drosophila* Ecd (22) showed Ecd to co-IP with U5 snRNA but the direct interaction of Ecd with the SmD3 component of the U5 Sm Ring implied that co-IP of Ecd and U5 snRNA may be indirect. Our results demonstrate the direct ECD binding to U5 snRNA and regulation of U5 snRNP complex (PRPF8, SNRNP200 and EFTUD2), supporting a model in which the binding of ECD to U5 snRNA is critical to help stabilize a fully assembled and functional U5 snRNP complex. ECD's RNA-independent association with the U5 snRNP protein components **(Fig. S5)**, along with concurrent binding to its RNA component, U5 snRNA, suggests that ECD is a major regulator of U5 during

its biogenesis and during the splicing process, potentially serving as a molecular glue to help U5 assembly and stability. Consistent with this model, mutations or loss of U5 snRNA in yeast significantly reduce the Prp8 protein levels emphasizing the interdependence between snRNAs and spliceosome proteins (64). In addition, ECD may also help recruit key additional proteins to help assemble and/or stabilize the U5 snRNP. For example, ECD is well-established to form a protein complex with the R2TP co-chaperone (10). Notably, PRPF8 was shown to assemble into a complex that included ECD, R2TP and additional proteins including AAR2 (65). AAR2 itself is a critical factor in U5 snRNP assembly (66). Pertinent to ECD's dual interaction with RNA and protein components of complex molecular machines like the U5 snRNP, it will be of great interest to explore if ECD plays a role in the assembly of other RNA-protein complexes, such as the small nucleolar ribonucleoprotein (snoRNP) complexes, which are well-established to require the ECD interacting R2TP complex for their assembly (67).

While our studies are consistent with a role of ECD in the cytoplasmic assembly of U5 snRNP, as shown for *Drosophila* Ecd (22), the direct binding of ECD to a large repertoire of mRNAs in exonic regions near their splice junctions supports its role in splicing beyond the U5 snRNP assembly. In this regard, it is notable that we have found ECD to shuttle between the cytoplasm and nucleus (68). Whether ECD shuttling might represent its necessity to maintain an intact U5 snRNP complex even during the steps of RNA splicing in the nucleus, is an important question that needs further exploration.

In addition to the mechanistic insights that the RNA binding ability of ECD provides to understand its roles in subcellular processes such as RNA splicing, our results provide convincing evidence that ECD's new role as an RBP is required for its cellular function. In a functional assay

with cell proliferation as a readout, the RNA binding deficient mutant of ECD failed to rescue MEF cells from a proliferative block imposed by the endogenous Ecd KO **(Fig. 8B-D).**

Taken together, our findings elucidate novel aspects of the mechanistic basis for ECD's role in maintaining the fidelity and efficiency of pre-mRNA splicing. Given the critical physiological functions of ECD and the pro-oncogenic effects due to its overexpression in cancers, these new mechanistic insights add a novel perspective in understanding its roles in cellular and cancer biology and devising means to disable ECD's function for therapeutic purposes.

Data availability:

The e-CLIP and RNA sequencing data generated in this study is deposited under Gene Expression Omnibus (GEO) accession: GSE285409.

Acknowledgements:

This research was funded by Pilot grants from the Fred & Pamela Buffett Cancer Center (HB & VB); Department of Defense grants W81XWH-17-1-0616 and W81XWH-20-1-0058 to HB and W81XWH-20-1-0546; HT94252410337; and National Institutes of Health (NIH) grants R21CA241055 and R03CA253193 to VB; the NIH MIRA award R35 GM147467 (to MJR); the Raphael Bonita Memorial Fund and support to UNMC core facilities from the NCI Cancer Center Support Grant (P30CA036727) awarded to Fred & Pamela Buffett Cancer Center and from the Nebraska Research Initiative.

Author contributions:

Conceptualization: MR, HB, MJR, and VB. Methodology, Data Analysis and Interpretation: MR, ARR, AK, IS, BBK, KKB, HB, MJR, and VB. Writing – original draft: MR. Writing – review and editing: MR, ARR, AK, IS, BBK, KKB, HB, MJR, and VB. Bioinformatic analyses: MR, AK and MJR. Supervision: HB, MJR, and VB. Funding acquisition: HB, MJR, and VB.

Supplementary Data Statement:

Supplementary Data are available at NAR online.

Conflict of Interest:

Dr. H. Band and Dr. V. Band received funding from Nimbus Therapeutics for an unrelated project. Other authors declare no conflict of interest.

References:

1. Muller-McNicoll, M. and Neugebauer, K.M. (2013) How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nat Rev Genet*, **14**, 275-287.
2. Gerstberger, S., Hafner, M. and Tuschl, T. (2014) A census of human RNA-binding proteins. *Nat Rev Genet*, **15**, 829-845.
3. Corley, M., Burns, M.C. and Yeo, G.W. (2020) How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Mol Cell*, **78**, 9-29.
4. Avila-Lopez, P. and Lauberth, S.M. (2024) Exploring new roles for RNA-binding proteins in epigenetic and gene regulation. *Curr Opin Genet Dev*, **84**, 102136.
5. Xu, Y., Huangyang, P., Wang, Y., Xue, L., Devericks, E., Nguyen, H.G., Yu, X., Oses-Prieto, J.A., Burlingame, A.L., Miglani, S. *et al.* (2021) ERalpha is an RNA-binding protein sustaining tumor cell survival and drug resistance. *Cell*, **184**, 5215-5229 e5217.
6. Beltran, M., Yates, C.M., Skalska, L., Dawson, M., Reis, F.P., Viiri, K., Fisher, C.L., Sibley, C.R., Foster, B.M., Bartke, T. *et al.* (2016) The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res*, **26**, 896-907.
7. Kanakkanthara, A., Jeganathan, K.B., Limzerwala, J.F., Baker, D.J., Hamada, M., Nam, H.J., van Deursen, W.H., Hamada, N., Naylor, R.M., Becker, N.A. *et al.* (2016) Cyclin A2 is an RNA binding protein that controls Mre11 mRNA translation. *Science*, **353**, 1549-1552.
8. Zhang, Y., Chen, J., Gurumurthy, C.B., Kim, J., Bhat, I., Gao, Q., Dimri, G., Lee, S.W., Band, H. and Band, V. (2006) The human orthologue of Drosophila ecdysoneless protein interacts with p53 and regulates its function. *Cancer Res*, **66**, 7167-7175.
9. Kim, J.H., Gurumurthy, C.B., Naramura, M., Zhang, Y., Dudley, A.T., Doglio, L., Band, H. and Band, V. (2009) Role of mammalian Ecdysoneless in cell cycle regulation. *J Biol Chem*, **284**, 26402-26410.
10. Mir, R.A., Bele, A., Mirza, S., Srivastava, S., Olou, A.A., Ammons, S.A., Kim, J.H., Gurumurthy, C.B., Qiu, F., Band, H. *et al.* (2015) A Novel Interaction of Ecdysoneless (ECD) Protein with R2TP Complex Component RUVBL1 Is Required for the Functional Role of ECD in Cell Cycle Progression. *Mol Cell Biol*, **36**, 886-899.
11. Olou, A.A., Sarkar, A., Bele, A., Gurumurthy, C.B., Mir, R.A., Ammons, S.A., Mirza, S., Saleem, I., Urano, F., Band, H. *et al.* (2017) Mammalian ECD Protein Is a Novel Negative Regulator of the PERK Arm of the Unfolded Protein Response. *Mol Cell Biol*, **37**.
12. Bele, A., Mirza, S., Zhang, Y., Ahmad Mir, R., Lin, S., Kim, J.H., Gurumurthy, C.B., West, W., Qiu, F., Band, H. *et al.* (2015) The cell cycle regulator ecdysoneless cooperates with H-Ras to promote oncogenic transformation of human mammary epithelial cells. *Cell Cycle*, **14**, 990-1000.
13. Mirza, S., Kalluchi, A., Raza, M., Saleem, I., Mohapatra, B., Pal, D., Ouellette, M.M., Qiu, F., Yu, L., Lobanov, A. *et al.* (2022) Ecdysoneless Protein Regulates Viral and Cellular mRNA Splicing to Promote Cervical Oncogenesis. *Mol Cancer Res*, **20**, 305-318.
14. Mohapatra, B.C., Mirza, S., Bele, A., Gurumurthy, C.B., Raza, M., Saleem, I., Storck, M.D., Sarkar, A., Kollala, S.S., Shukla, S.K. *et al.* (2022) Ecdysoneless Overexpression

Drives Mammary Tumorigenesis through Upregulation of C-MYC and Glucose Metabolism. *Mol Cancer Res*, **20**, 1391-1404.
15. Xu, S.H., Zhu, S., Wang, Y., Huang, J.Z., Chen, M., Wu, Q.X., He, Y.T., Chen, D. and Yan, G.R. (2018) ECD promotes gastric cancer metastasis by blocking E3 ligase ZFP91-mediated hnRNP F ubiquitination and degradation. *Cell Death Dis*, **9**, 479.
16. Dey, P., Rachagani, S., Chakraborty, S., Singh, P.K., Zhao, X., Gurumurthy, C.B., Anderson, J.M., Lele, S., Hollingsworth, M.A., Band, V. *et al.* (2012) Overexpression of ecdysoneless in pancreatic cancer and its role in oncogenesis by regulating glycolysis. *Clin Cancer Res*, **18**, 6188-6198.
17. Xu, S.H., Huang, J.Z., Xu, M.L., Yu, G., Yin, X.F., Chen, D. and Yan, G.R. (2015) ACK1 promotes gastric cancer epithelial-mesenchymal transition and metastasis through AKT-POU2F1-ECD signalling. *J Pathol*, **236**, 175-185.
18. Zhao, X., Mirza, S., Alshareeda, A., Zhang, Y., Gurumurthy, C.B., Bele, A., Kim, J.H., Mohibi, S., Goswami, M., Lele, S.M. *et al.* (2012) Overexpression of a novel cell cycle regulator ecdysoneless in breast cancer: a marker of poor prognosis in HER2/neu-overexpressing breast cancer patients. *Breast Cancer Res Treat*, **134**, 171-180.
19. Saleem, I., Mirza, S., Sarkar, A., Raza, M., Mohapatra, B., Mushtaq, I., Kim, J.H., Mishra, N.K., Alsaleem, M.A., Rakha, E.A. *et al.* (2021) The Mammalian Ecdysoneless Protein Interacts with RNA Helicase DDX39A To Regulate Nuclear mRNA Export. *Mol Cell Biol*, **41**, e0010321.
20. Gaziova, I., Bonnette, P.C., Henrich, V.C. and Jindra, M. (2004) Cell-autonomous roles of the ecdysoneless gene in Drosophila development and oogenesis. *Development*, **131**, 2715-2725.
21. Claudius, A.K., Romani, P., Lamkemeyer, T., Jindra, M. and Uhlirova, M. (2014) Unexpected role of the steroid-deficiency protein ecdysoneless in pre-mRNA splicing. *PLoS Genet*, **10**, e1004287.
22. Erkelenz, S., Stankovic, D., Mundorf, J., Bresser, T., Claudius, A.K., Boehm, V., Gehring, N.H. and Uhlirova, M. (2021) Ecd promotes U5 snRNP maturation and Prp8 stability. *Nucleic Acids Res*, **49**, 1688-1707.
23. Mir, R.A., Lovelace, J., Schafer, N.P., Simone, P.D., Kellezi, A., Kolar, C., Spagnol, G., Sorgen, P.L., Band, H., Band, V. *et al.* (2016) Biophysical characterization and modeling of human Ecdysoneless (ECD) protein supports a scaffolding function. *AIMS Biophys*, **3**, 195-208.
24. Band, V. and Sager, R. (1989) Distinctive traits of normal and tumor-derived human mammary epithelial cells expressed in a medium that supports long-term growth of both cell types. *Proc Natl Acad Sci U S A*, **86**, 1249-1253.
25. Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J. *et al.* (2024) AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res*, **52**, D368-D375.
26. Van Nostrand, E.L., Nguyen, T.B., Gelboin-Burkhart, C., Wang, R., Blue, S.M., Pratt, G.A., Louie, A.L. and Yeo, G.W. (2017) Robust, Cost-Effective Profiling of RNA Binding Protein Targets with Single-end Enhanced Crosslinking and Immunoprecipitation (seCLIP). *Methods Mol Biol*, **1648**, 177-200.

27. Kivioja, T., Vaharautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S. and Taipale, J. (2011) Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods*, **9**, 72-74.
28. Smith, T., Heger, A. and Sudbery, I. (2017) UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res*, **27**, 491-499.
29. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011*, **17**, 3.
30. Hubley, R., Finn, R.D., Clements, J., Eddy, S.R., Jones, T.A., Bao, W., Smit, A.F. and Wheeler, T.J. (2016) The Dfam database of repetitive DNA families. *Nucleic Acids Res*, **44**, D81-89.
31. Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Sayers, E.W. (2013) GenBank. *Nucleic Acids Res*, **41**, D36-42.
32. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15-21.
33. Lovci, M.T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T.Y., Stark, T.J., Gehman, L.T., Hoon, S. *et al.* (2013) Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat Struct Mol Biol*, **20**, 1434-1442.
34. Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J. *et al.* (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*, **47**, D766-D773.
35. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, **38**, 576-589.
36. Mahony, S. and Benos, P.V. (2007) STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic Acids Res*, **35**, W253-258.
37. Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.Y., Cody, N.A.L., Dominguez, D. *et al.* (2020) A large-scale binding and functional map of human RNA-binding proteins. *Nature*, **583**, 711-719.
38. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.
39. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T. and Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*, **33**, 290-295.
40. Wang, Y., Xie, Z., Kutschera, E., Adams, J.I., Kadash-Edmondson, K.E. and Xing, Y. (2024) rMATS-turbo: an efficient and flexible computational tool for alternative splicing analysis of large-scale RNA-seq data. *Nat Protoc*, **19**, 1083-1104.
41. Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M. *et al.* (2021) Gene Set Knowledge Discovery with Enrichr. *Curr Protoc*, **1**, e90.
42. Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A. and Manke, T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*, **42**, W187-191.
43. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat Biotechnol*, **29**, 24-26.

44. Lachmann, A., Torre, D., Keenan, A.B., Jagodnik, K.M., Lee, H.J., Wang, L., Silverstein, M.C. and Ma'ayan, A. (2018) Massive mining of publicly available RNA-seq data from human and mouse. *Nat Commun*, **9**, 1366.
45. Terribilini, M., Sander, J.D., Lee, J.H., Zaback, P., Jernigan, R.L., Honavar, V. and Dobbs, D. (2007) RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. *Nucleic Acids Res*, **35**, W578-584.
46. Sugiura, T., Nagano, Y. and Noguchi, Y. (2007) DDX39, upregulated in lung squamous cell cancer, displays RNA helicase activities and promotes cancer cell growth. *Cancer Biol Ther*, **6**, 957-964.
47. Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K. *et al.* (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods*, **13**, 508-514.
48. Glisovic, T., Bachorik, J.L., Yong, J. and Dreyfuss, G. (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett*, **582**, 1977-1986.
49. Peng, Q., Zhou, Y., Oyang, L., Wu, N., Tang, Y., Su, M., Luo, X., Wang, Y., Sheng, X., Ma, J. *et al.* (2022) Impacts and mechanisms of alternative mRNA splicing in cancer metabolism, immune response, and therapeutics. *Mol Ther*, **30**, 1018-1035.
50. Tao, Y., Zhang, Q., Wang, H., Yang, X. and Mu, H. (2024) Alternative splicing and related RNA binding proteins in human health and disease. *Signal Transduct Target Ther*, **9**, 26.
51. Dreyfuss, G., Kim, V.N. and Kataoka, N. (2002) Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol*, **3**, 195-205.
52. Li, W., Deng, X. and Chen, J. (2022) RNA-binding proteins in regulating mRNA stability and translation: roles and mechanisms in cancer. *Semin Cancer Biol*, **86**, 664-677.
53. Basu, S. and Bahadur, R.P. (2016) A structural perspective of RNA recognition by intrinsically disordered proteins. *Cell Mol Life Sci*, **73**, 4075-4084.
54. Tompa, P. and Csermely, P. (2004) The role of structural disorder in the function of RNA and protein chaperones. *The FASEB Journal*, **18**, 1169-1175.
55. Sarni, S.H., Roca, J., Du, C., Jia, M., Li, H., Damjanovic, A., Malecka, E.M., Wysocki, V.H. and Woodson, S.A. (2022) Intrinsically disordered interaction network in an RNA chaperone revealed by native mass spectrometry. *Proc Natl Acad Sci U S A*, **119**, e2208780119.
56. Lunde, B.M., Moore, C. and Varani, G. (2007) RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*, **8**, 479-490.
57. Watakabe, A., Tanaka, K. and Shimura, Y. (1993) The role of exon sequences in splice site selection. *Genes Dev*, **7**, 407-418.
58. Tacke, R. and Manley, J.L. (1999) Determinants of SR protein specificity. *Curr Opin Cell Biol*, **11**, 358-362.
59. Zheng, Z.-M. (2004) Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. *Journal of biomedical science*, **11**, 278-294.
60. Jarvelin, A.I., Noerenberg, M., Davis, I. and Castello, A. (2016) The new (dis)order in RNA regulation. *Cell Commun Signal*, **14**, 9.
61. Newman, A.J. (1997) The role of U5 snRNP in pre-mRNA splicing. *EMBO J*, **16**, 5797-5800.

62. Hacker, I., Sander, B., Golas, M.M., Wolf, E., Karagoz, E., Kastner, B., Stark, H., Fabrizio, P. and Luhrmann, R. (2008) Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy. *Nat Struct Mol Biol*, **15**, 1206-1212.
63. Ehsani, A., Alluin, J.V. and Rossi, J.J. (2013) Cell cycle abnormalities associated with differential perturbations of the human U5 snRNP associated U5-200kD RNA helicase. *PLoS One*, **8**, e62125.
64. Kershaw, C.J., Barrass, J.D., Beggs, J.D. and O'Keefe, R.T. (2009) Mutations in the U5 snRNA result in altered splicing of subsets of pre-mRNAs and reduced stability of Prp8. *RNA*, **15**, 1292-1304.
65. Serna, M., Gonzalez-Corpas, A., Cabezudo, S., Lopez-Perrote, A., Degliesposti, G., Zarzuela, E., Skehel, J.M., Munoz, J. and Llorca, O. (2022) CryoEM of RUVBL1-RUVBL2-ZNHIT2, a complex that interacts with pre-mRNA-processing-splicing factor 8. *Nucleic Acids Res*, **50**, 1128-1146.
66. Nguyen, T.H., Li, J., Galej, W.P., Oshikane, H., Newman, A.J. and Nagai, K. (2013) Structural basis of Brr2-Prp8 interactions and implications for U5 snRNP biogenesis and the spliceosome active site. *Structure*, **21**, 910-919.
67. Massenet, S., Bertrand, E. and Verheggen, C. (2017) Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA Biol*, **14**, 680-692.
68. Kim, J.H., Gurumurthy, C.B., Band, H. and Band, V. (2010) Biochemical characterization of human Ecdysoneless reveals a role in transcriptional regulation. *Biol Chem*, **391**, 9-19.

Figure 1

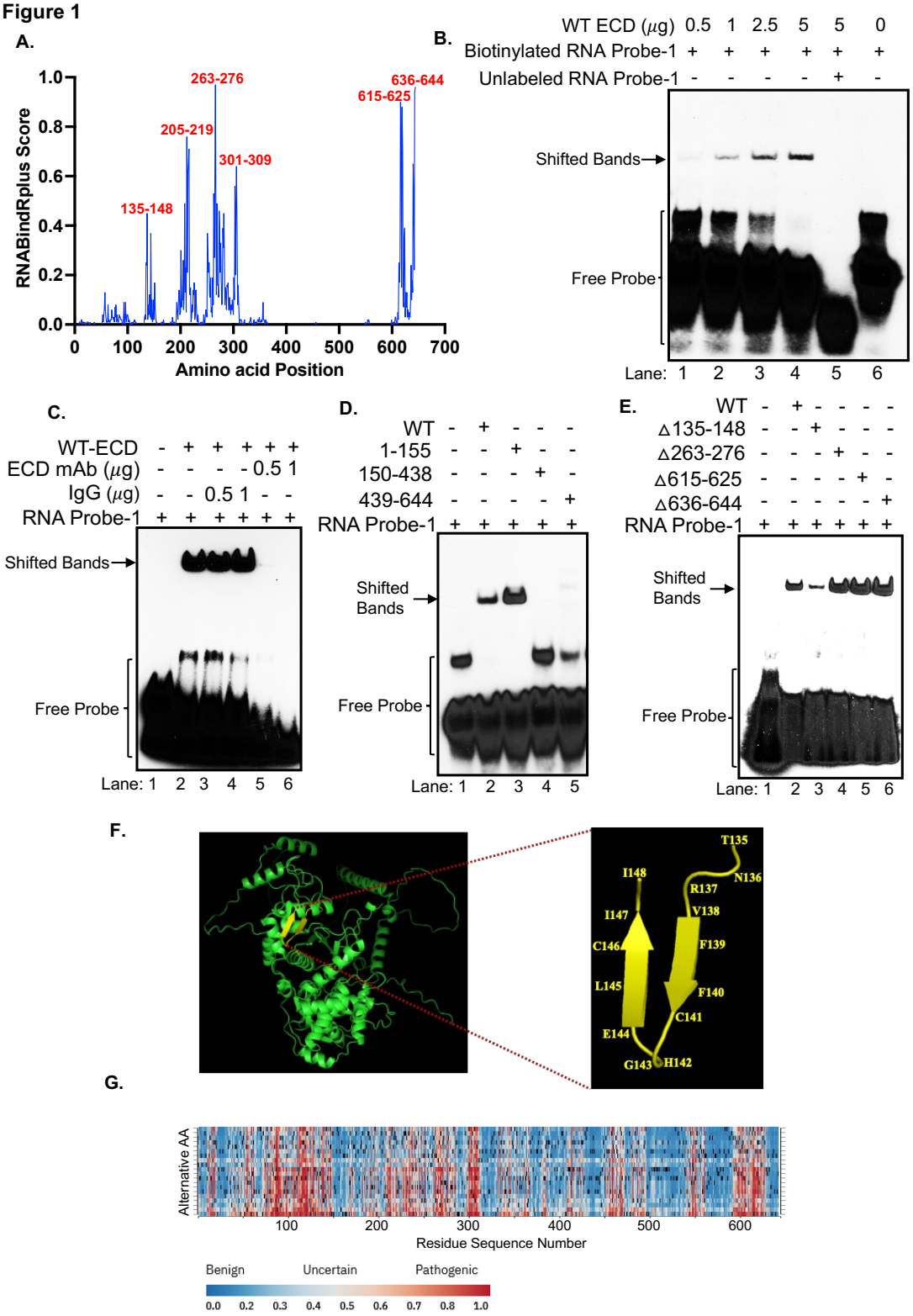

A.
RNABindRplus Score
Amino acid Position
135-148
205-219
263-276
301-309
615-625
636-644
B.
WT ECD (μg) 0.5 1 2.5 5 5 0
Biotinylated RNA Probe-1 + + + + + +
Unlabeled RNA Probe-1 - - - - + -
Shifted Bands
Free Probe
Lane: 1 2 3 4 5 6
C.
WT-ECD - + + + + +
ECD mAb (μg) - - - - 0.5 1
IgG (μg) - - 0.5 1 - -
RNA Probe-1 + + + + + +
Shifted Bands
Free Probe
Lane: 1 2 3 4 5 6
D.
WT - + - - -
1-155 - - + - -
150-438 - - - + -
439-644 - - - - +
RNA Probe-1 + + + + +
Shifted Bands
Free Probe
Lane: 1 2 3 4 5
E.
WT - + - - - -
Δ135-148 - - + - - -
Δ263-276 - - - + - -
Δ615-625 - - - - + -
Δ636-644 - - - - - +
RNA Probe-1 + + + + + +
Shifted Bands
Free Probe
Lane: 1 2 3 4 5 6
F.
T135
N136
I148
R137
V138
I147
C146
F139
L145
F140
C141
E144
G143
H142
G.
Alternative AA
Residue Sequence Number
Benign
Uncertain
Pathogenic

Fig. 1. ECD directly binds to RNA via the N-terminal region. (A) ECD amino acids predicted to bind RNA using online tool RNABindRplus**. (B)** RNA-EMSA with different concentrations of purified ECD (lanes 1 to 4) along with biotinylated RNA probe (40 nM) previously reported to bind DDX39A (Probe-1). Protein free probe, used as a negative control (lane 6). Competition with 100-fold excess unlabeled probe was done to determine the binding specificity (lane 5). **(C)** RNA EMSA with purified ECD (2.5µg), with RNA Probe-1 (all lanes), in the absence (lane 2) or presence of two different concentrations of mouse ECD monoclonal antibody (lanes 5 and 6) or equal amount of mouse IgG (lanes 3 and 4). Protein free probe served as negative control (lane 1). **(D)** RNA-EMSA with 2.5µg of wild type (WT, lane 1) or same amount of (2.5µg) indicated truncated mutants of ECD (lanes 2 to 5) along with Probe-1. **(E)** Purified deletion mutants used in RNA EMSA (2.5µg) show loss of RNA binding upon deletion of amino acids (aa) 135 to 148 (lane 3) while WT (lane 1) and other deletion mutants of ECD (lanes 4 to 6) show binding to RNA. All EMSA samples were run on 7.5% native gel, transferred to H-bond membrane, and developed with Streptavidin-HRP based chemiluminescent method. **(F)** Predicted tertiary structures of human ECD, showing identified RNA binding region, aa 135-148 (in yellow). PDB file of Alpha Fold predicted 3D structure of human ECD was visualized using PyMOL software. **(G)** Alpha Misfold prediction of mutation pathogenicity per residue of human ECD.

Figure 2

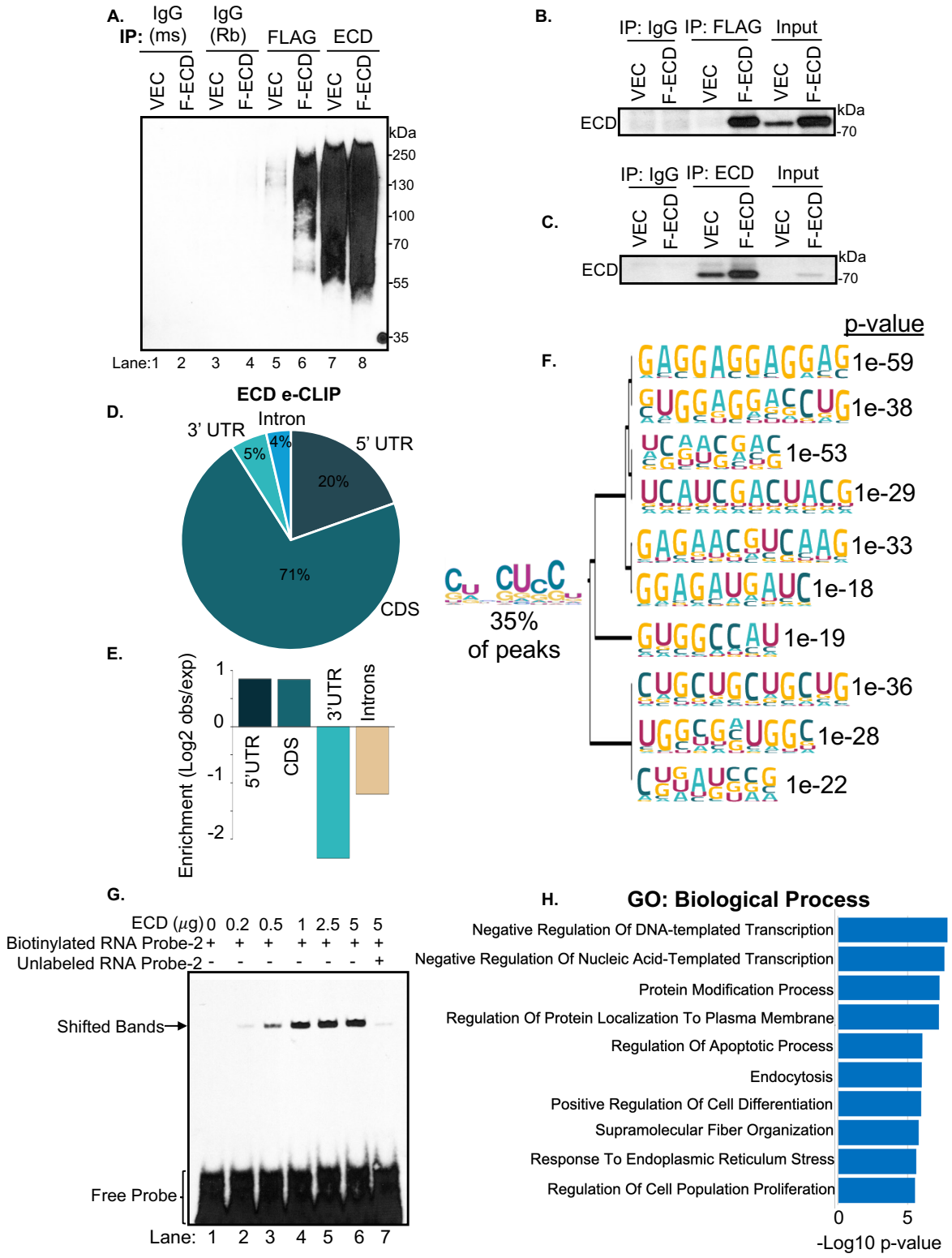

A.
IP: IgG (ms) IgG (Rb) FLAG ECD
VEC F-ECD VEC F-ECD VEC F-ECD VEC F-ECD
kDa
250
130
100
70
55
35
Lane:1 2 3 4 5 6 7 8
B.
IP: IgG IP: FLAG Input
VEC F-ECD VEC F-ECD VEC F-ECD
ECD
kDa
70
C.
IP: IgG IP: ECD Input
VEC F-ECD VEC F-ECD VEC F-ECD
ECD
kDa
70
D.
ECD e-CLIP
3' UTR
Intron
5' UTR
5%
4%
20%
71%
CDS
E.
Enrichment (Log2 obs/exp)
1
0
-1
-2
5'UTR
CDS
3'UTR
Introns
F.
p-value
GAGGAGGAGGAG 1e-59
GUGGAGGACCUG 1e-38
UCAACGAC 1e-53
UCAUCGACUACG 1e-29
GAGAACGUCAAG 1e-33
GGAGAUGAUC 1e-18
GUGGCCAU 1e-19
CUGCUGCUGCUG 1e-36
UGGCGGUGGC 1e-28
CUGAUCCG 1e-22
35%
of peaks
G.
ECD (μg) 0 0.2 0.5 1 2.5 5 5
Biotinylated RNA Probe-2 + + + + + + +
Unlabeled RNA Probe-2 - - - - - - +
Shifted Bands
Free Probe
Lane: 1 2 3 4 5 6 7
H.
GO: Biological Process
Negative Regulation Of DNA-templated Transcription
Negative Regulation Of Nucleic Acid-Templated Transcription
Protein Modification Process
Regulation Of Protein Localization To Plasma Membrane
Regulation Of Apoptotic Process
Endocytosis
Positive Regulation Of Cell Differentiation
Supramolecular Fiber Organization
Response To Endoplasmic Reticulum Stress
Regulation Of Cell Population Proliferation
0
5
-Log10 p-value

Fig. 2. RNA binding landscape of ECD by eCLIP seq. (A) Blot depicting immunoprecipitated (IP) RNA. Protein lysates from UV crosslinked 76NTERT cells expressing vector or FLAG-ECD were subjected to IP using 5 µg each of anti-ECD or FLAG or respective IgG (rabbit-Rb) and IgG (mouse-ms) antibodies. The transcripts pull down by indicated antibodies were end repaired and ligated with a biotin adapter at 3' end as per the protocol by ECLIPSEBIO antibody IP validation kit. IP samples were run on NuPAGE 4-12% Bis-Tris protein gels, transferred to nitrocellulose membrane, and visualized using the chemiluminescent nucleic acid detection method. **(B&C)** WB depicting IPed ECD protein using anti-FLAG **(B)** and anti-ECD **(C)** antibodies. For WB, IP and 3% of input samples were run on SDS-PAGE gels and transferred to PVDF membrane and probed with ECD antibody. **(D)** The distribution profile of the ECD eCLIP peaks in the human genome from biological duplicates. CDS-coding sequence, UTR-untranslated region. **(E)** Enrichment of ECD eCLIP-seq peaks compared to random distribution (expected) to account for the different frequency and size of individual elements. **(F)** The top-ranked motifs discovered de novo from ECD eCLIP-seq peaks by HOMER analysis along with a consensus motif (present in 35% of ECD eCLIP peaks) identified through clustering of top 10 HOMER motifs using STAMP. **(G)** RNA-EMSA with indicated concentrations of purified ECD (lanes 2 to 4) along with 40 nM of eCLIP identified motif used as biotinylated RNA probe (Probe-2). Protein free probe, used as a negative control (lane 1). Competition with 100-fold excess unlabeled probe was performed to determine the binding specificity (lane 7). Samples were processed as mentioned in Fig. 1. **(H)** Top pathways affected by genes harboring ECD binding consensus motif as identified through GO Biological pathway analysis.

Figure 3

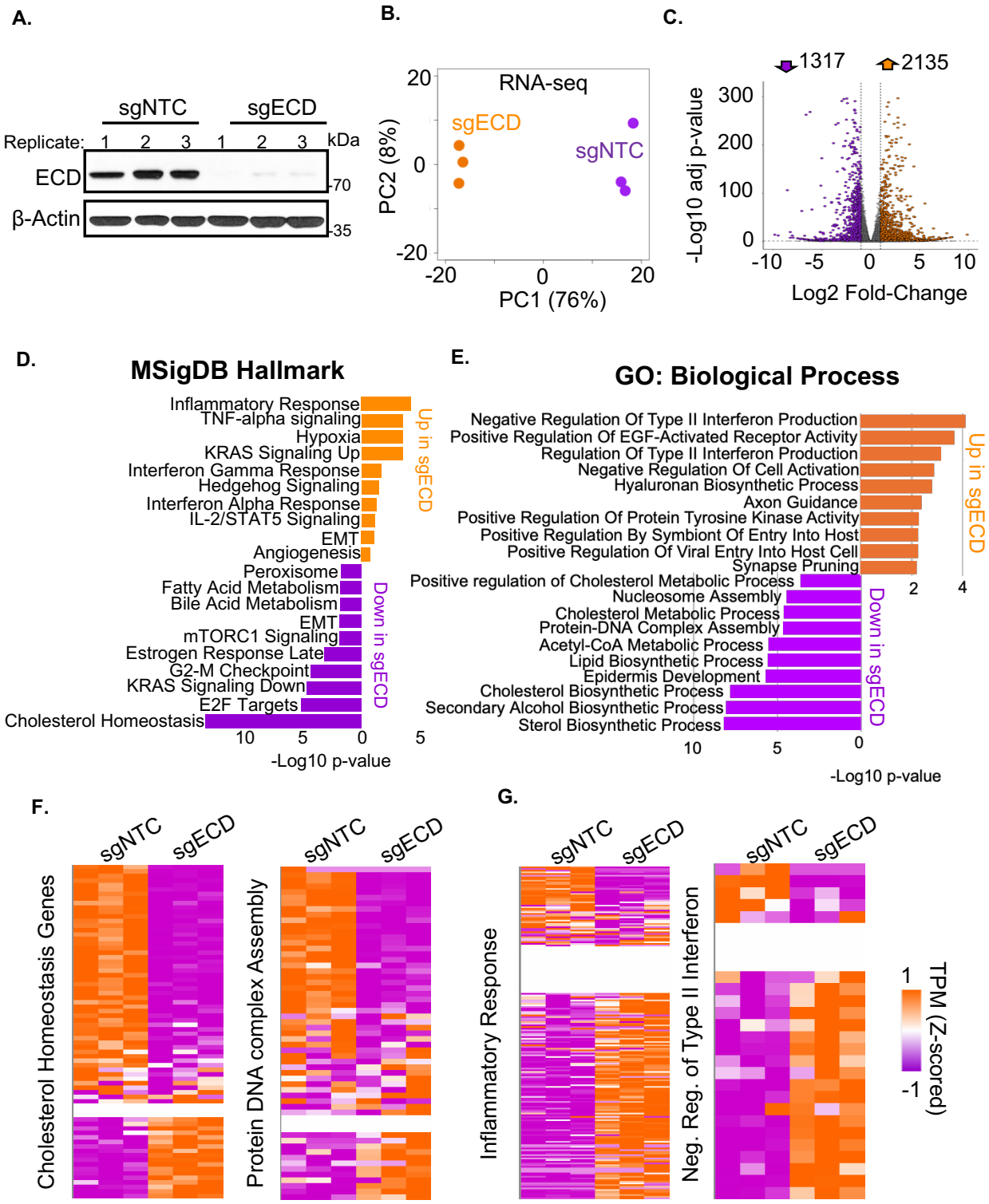

A.
sgNTC
sgECD
Replicate: 1 2 3 1 2 3 kDa
ECD
β-Actin
70
35
B.
RNA-seq
sgECD
sgNTC
PC2 (8%)
PC1 (76%)
20
0
-20
-20
0
20
C.
1317
2135
-Log10 adj p-value
Log2 Fold-Change
300
200
100
0
-10
-5
0
5
10
D.
MSigDB Hallmark
Inflammatory Response
TNF-alpha signaling
Hypoxia
KRAS Signaling Up
Interferon Gamma Response
Hedgehog Signaling
Interferon Alpha Response
IL-2/STAT5 Signaling
EMT
Angiogenesis
Peroxisome
Fatty Acid Metabolism
Bile Acid Metabolism
EMT
mTORC1 Signaling
Estrogen Response Late
G2-M Checkpoint
KRAS Signaling Down
E2F Targets
Cholesterol Homeostasis
Up in sgECD
Down in sgECD
10
5
0
5
-Log10 p-value
E.
GO: Biological Process
Negative Regulation Of Type II Interferon Production
Positive Regulation Of EGF-Activated Receptor Activity
Regulation Of Type II Interferon Production
Negative Regulation Of Cell Activation
Hyaluronan Biosynthetic Process
Axon Guidance
Positive Regulation Of Protein Tyrosine Kinase Activity
Positive Regulation By Symbiont Of Entry Into Host
Positive Regulation Of Viral Entry Into Host Cell
Synapse Pruning
Positive regulation of Cholesterol Metabolic Process
Nucleosome Assembly
Cholesterol Metabolic Process
Protein-DNA Complex Assembly
Acetyl-CoA Metabolic Process
Lipid Biosynthetic Process
Epidermis Development
Cholesterol Biosynthetic Process
Secondary Alcohol Biosynthetic Process
Sterol Biosynthetic Process
Up in sgECD
Down in sgECD
2
4
10
5
0
-Log10 p-value
F.
sgNTC
sgECD
Cholesterol Homeostasis Genes
Protein DNA complex Assembly
G.
Inflammatory Response
Neg. Reg. of Type II Interferon
TPM (Z-scored)
1
-1

Fig. 3. ECD depletion alters transcriptomic profile. (A) WB depicting expression of ECD protein in the control (sgNTC) and ECD depleted (sgECD) 76NTERT cells used for RNA seq analysis, following doxycycline-induced Cas9 activation. β-Actin used as loading control. **(B)** PCA of RNA-seq data done in biological triplicates after ECD depletion (orange) compared to non-targeting controls (sgNTC, purple). **(C)** Volcano plot of differential expression after ECD depletion identified thousands (1,317 decreased and 2,135 increased) of significant differentially expressed genes (Benjamini-Hochberd corrected p-value < .05 and >2-fold change). **(D-E)** The top 10 increased and decreased MsigDB Hallmark pathways and Gene Ontology Molecular Function gene sets associated with differential expression post ECD depletion. **(F-G)** Heatmaps of the z-scored expression (Transcripts Per Million) for the specified gene sets.

Figure 4

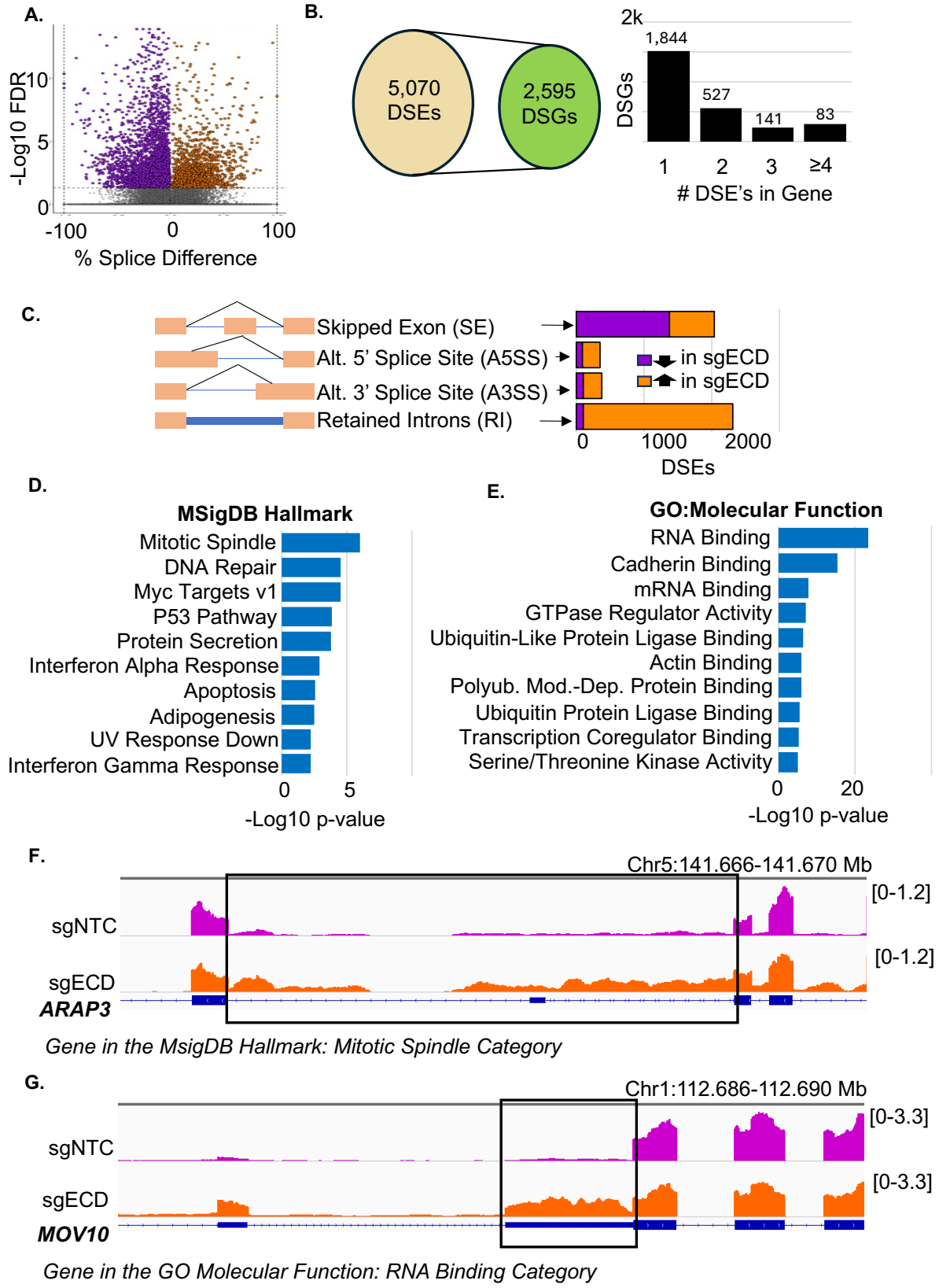

A.
-Log10 FDR
10
5
0
-100
0
100
% Splice Difference
B.
5,070 DSEs
2,595 DSGs
2k
DSGs
1,844
527
141
83
1
2
3
≥4
DSE's in Gene
C.
Skipped Exon (SE)
Alt. 5' Splice Site (A5SS)
Alt. 3' Splice Site (A3SS)
Retained Introns (RI)
in sgECD
in sgECD
0
1000
2000
DSEs
D.
MSigDB Hallmark
Mitotic Spindle
DNA Repair
Myc Targets v1
P53 Pathway
Protein Secretion
Interferon Alpha Response
Apoptosis
Adipogenesis
UV Response Down
Interferon Gamma Response
0
5
-Log10 p-value
E.
GO:Molecular Function
RNA Binding
Cadherin Binding
mRNA Binding
GTPase Regulator Activity
Ubiquitin-Like Protein Ligase Binding
Actin Binding
Polyub. Mod.-Dep. Protein Binding
Ubiquitin Protein Ligase Binding
Transcription Coregulator Binding
Serine/Threonine Kinase Activity
0
20
-Log10 p-value
F.
Chr5:141.666-141.670 Mb
sgNTC
[0-1.2]
sgECD
[0-1.2]
ARAP3
Gene in the MsigDB Hallmark: Mitotic Spindle Category
G.
Chr1:112.686-112.690 Mb
sgNTC
[0-3.3]
sgECD
[0-3.3]
MOV10
Gene in the GO Molecular Function: RNA Binding Category

Fig. 4. ECD expression impacts RNA splicing. (A) Results of differential splicing analysis, which identified thousands of significant (FDR < .05) differential splice events (DSEs). Schematic and Bar graph showing that **(B)** many genes have multiple differential splice events (e.g. multiple retained introns), such that 5,070 differential splice events are mapped to 2,595 genes and **(C)** categorization of differential splice events identified primarily aberrant increased retention of features (an increase of retained introns and a decrease of skipped exons). **(D-E)** The top MsigDB Hallmark pathways and Gene Ontology Molecular Function gene sets associated with differentially spliced genes upon ECD loss. **(F-G)** RNA-seq signal track showing an aberrant retention of a gene (in the top categories shown in D and E) post ECD depletion.

Figure 5

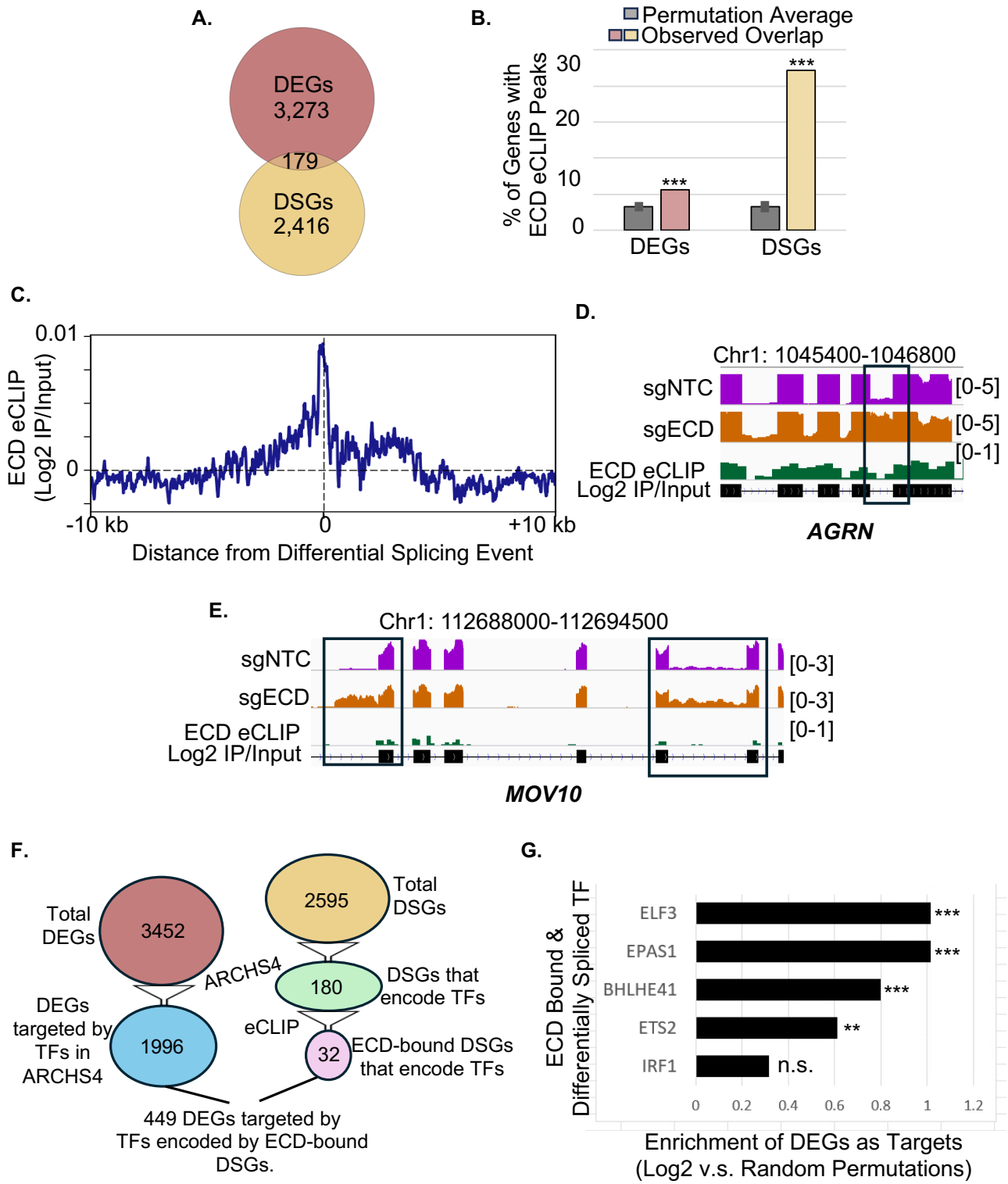

A.
DEGs
3,273
179
DSGs
2,416
B.
Permutation Average
Observed Overlap
% of Genes with
ECD eCLIP Peaks
30
20
10
0
***
***
DEGs
DSGs
C.
0.01
ECD eCLIP
(Log2 IP/Input)
0
-10 kb
0
+10 kb
Distance from Differential Splicing Event
D.
Chr1: 1045400-1046800
sgNTC
[0-5]
sgECD
[0-5]
[0-1]
ECD eCLIP
Log2 IP/Input
AGRN
E.
Chr1: 112688000-112694500
sgNTC
[0-3]
sgECD
[0-3]
[0-1]
ECD eCLIP
Log2 IP/Input
MOV10
F.
Total
DEGs
3452
2595
Total
DSGs
ARCHS4
180
DSGs that
encode TFs
DEGs
targeted by
TFs in
ARCHS4
1996
eCLIP
32
ECD-bound DSGs
that encode TFs
449 DEGs targeted by
TFs encoded by ECD-bound
DSGs.
G.
ECD Bound &
Differentially Spliced TF
ELF3
***
EPAS1
***
BHLHE41
***
ETS2
**
IRF1
n.s.
0
0.2
0.4
0.6
0.8
1
1.2
Enrichment of DEGs as Targets
(Log2 v.s. Random Permutations)

Fig. 5. ECD binds RNA to modulate splicing. (A) Venn diagram depicting, differentially expressed genes (DEGs) and differentially spliced genes (DSGs) have minimal overlap. **(B)** Bar graphs showing percentage of differentially expressed genes (DEGs) and differentially spliced genes (DSGs) that overlap an ECD e-CLIP peak. *** indicate p-value < .001 (Monte Carlo permutation testing) compared to the average overlap of peaks with randomly selected genes (grey) with error bars representing the standard deviation from a 1,000 permutation tests. **(C)** Average profile of ECD e-CLIP enrichment signal (Log2 IP/input) at differentially spliced coordinates. **(D & E)** Examples of RNA-seq signal at introns that are retained after ECD depletion. Shown below is the ECD e-CLIP signal in the control cells. **(F)** Schematic of the filtering pipeline to identify gene targets as ECD bound transcripts that are differential spliced and encode transcription factors (TFs) that putatively regulate differentially expressed genes in the ARCHS4 database. **(G)** Top 5 TFs based on the number of DEGs that they regulate relative to randomly selected genes. *** p<.001 ** p<.01 n.s. not significant, computed through Monte Carlo permutation testing.

Figure 6

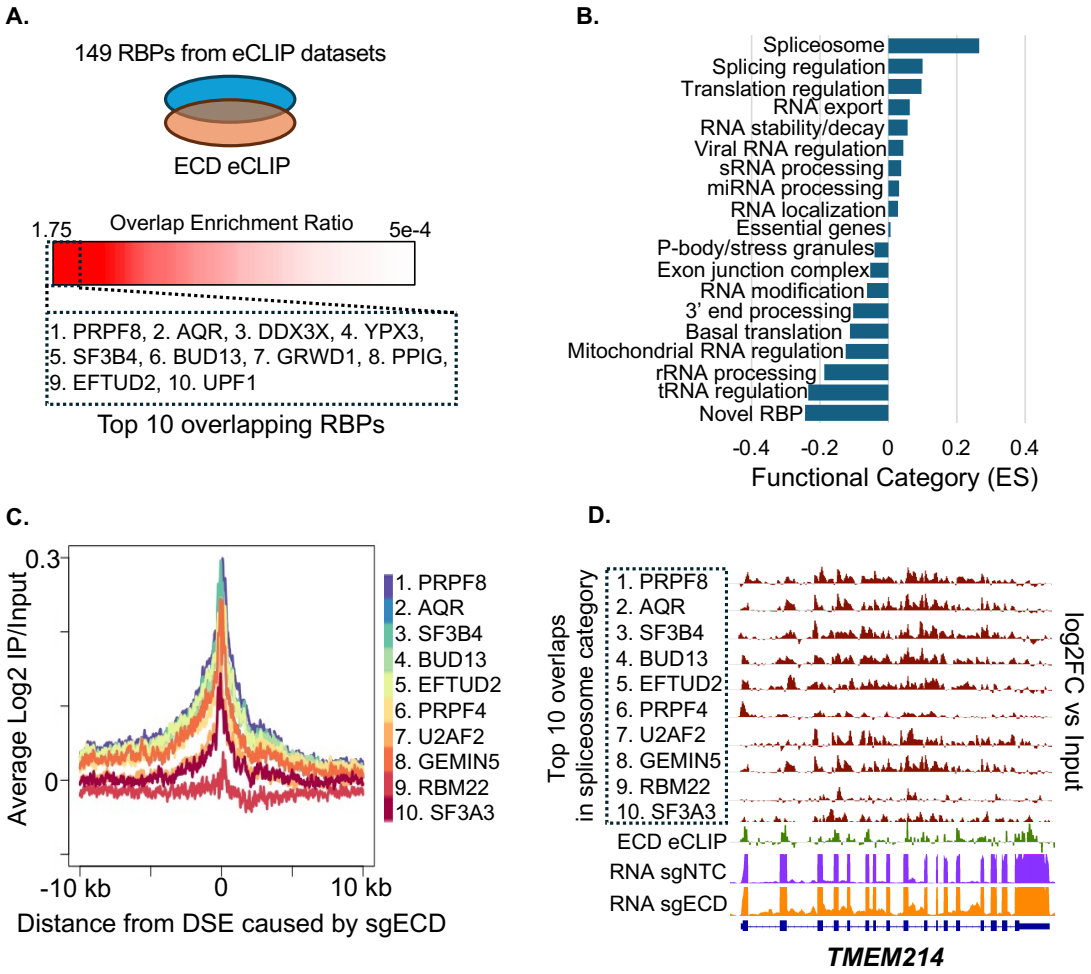


Fig. 6. Transcriptomic landscape bound by ECD overlaps with RNA binding proteins involved in splicing regulation. (A) 149 RBPs with eCLIP data were tested for overlap of targets with that of ECD eCLIP. Overlap enrichment ratio is calculated as the fraction of ECD peaks that overlap the RBP peaks vs those that don't. The top 10 RBPs (ranked by overlap enrichment) are listed. **(B)** The relative enrichment of functional categories associated with RBPs that overlap ECD eCLIP targets. (ES-enrichment score). **(C)** Average signal profiles of those top 10 spliceosome RBPs centered around differential splice events caused by sgECD. **(D)** eCLIP signals at an example gene, *TMEM214* for each of the top 10 RBPs in the spliceosome category and ECD.

Figure 7

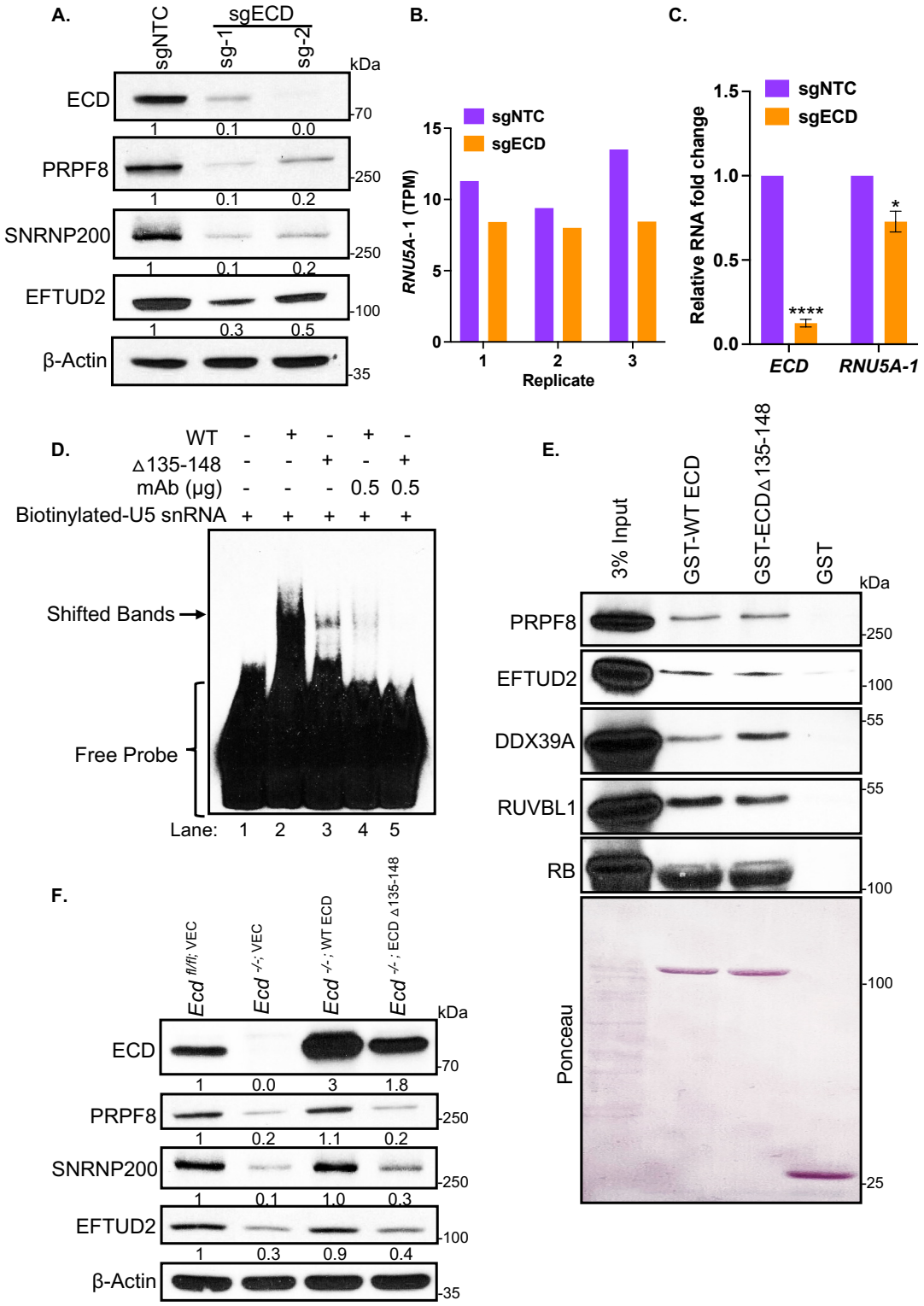

A.
sgNTC
sgECD
sg-1
sg-2
kDa
ECD
70
1 0.1 0.0
PRPF8
250
1 0.1 0.2
SNRNP200
250
1 0.1 0.2
EFTUD2
100
1 0.3 0.5
β-Actin
35
B.
sgNTC
sgECD
RNU5A-1 (TPM)
15
10
5
0
1
2
3
Replicate
C.
sgNTC
sgECD
Relative RNA fold change
1.5
1.0
0.5
0.0
****
*
ECD
RNU5A-1
D.
WT - + - + -
Δ135-148 - - + - +
mAb (μg) - - - 0.5 0.5
Biotinylated-U5 snRNA + + + + +
Shifted Bands
Free Probe
Lane: 1 2 3 4 5
E.
3% Input
GST-WT ECD
GST-ECDΔ135-148
GST
kDa
PRPF8
250
EFTUD2
100
DDX39A
55
RUVBL1
55
RB
100
Ponceau
100
25
F.
Ecd fl/fl; VEC
Ecd -/-; VEC
Ecd -/-; WT ECD
Ecd -/-; ECD Δ135-148
kDa
ECD
70
1 0.0 3 1.8
PRPF8
250
1 0.2 1.1 0.2
SNRNP200
250
1 0.1 1.0 0.3
EFTUD2
100
1 0.3 0.9 0.4
β-Actin
35

Fig. 7. ECD depletion results in loss of key components of U5 snRNP. (A) Western blot (WB) analysis of indicated protein levels in 76NTERT cells with NTC or ECD-targeted sgRNAs (sg-1/2) following doxycycline-induced Cas9 activation. Numbers below the blots depict band intensities in respect to loading control β-Actin, determined using Image J. **(B)** Histograms showing expression in TPM of *RNU5A-1* transcript in three biological replicates from control and ECD depleted cells as identified through RNA seq analysis. **(C)** Relative fold change in expression of *ECD* and *RNU5A-1* genes in control (sgNTC) and ECD depleted cells (sgECD), as determined through qPCR experiments.18s rRNA used as normalizing control. The data presented as mean ± SEM of three independent experiments. **** indicate p-value <.0001; * p-value <0.05, computed through unpaired student t test. **(D)** RNA-EMSA with U5 snRNA as probe along with 2.5μg each of WT or RBD mutant of ECD in the presence or absence of ECD mAb. Experiment was performed as indicated in Fig. 1. **(E)** WB show ECD and RBD mutant associated proteins using GST pull down assay. 20 μg of GST-tagged WT-ECD, its RBD mutant or GST alone (negative control) immobilized on glutathione-Sepharose beads were incubated with 1 mg of protein lysate from 76NTERT and associated proteins were analyzed by WB with the indicated antibodies. The membrane was stained with Ponceau dye to visualize the quantity of fusion proteins used for pulldowns. **(F)** WB shows expression of indicated proteins in control or *Ecd* KO MEFs and upon rescue with either WT human ECD or RBD (Δ135-148) mutant. Numbers below the blots depict band intensities in respect to loading control β-Actin, determined using Image J.

Figure 8

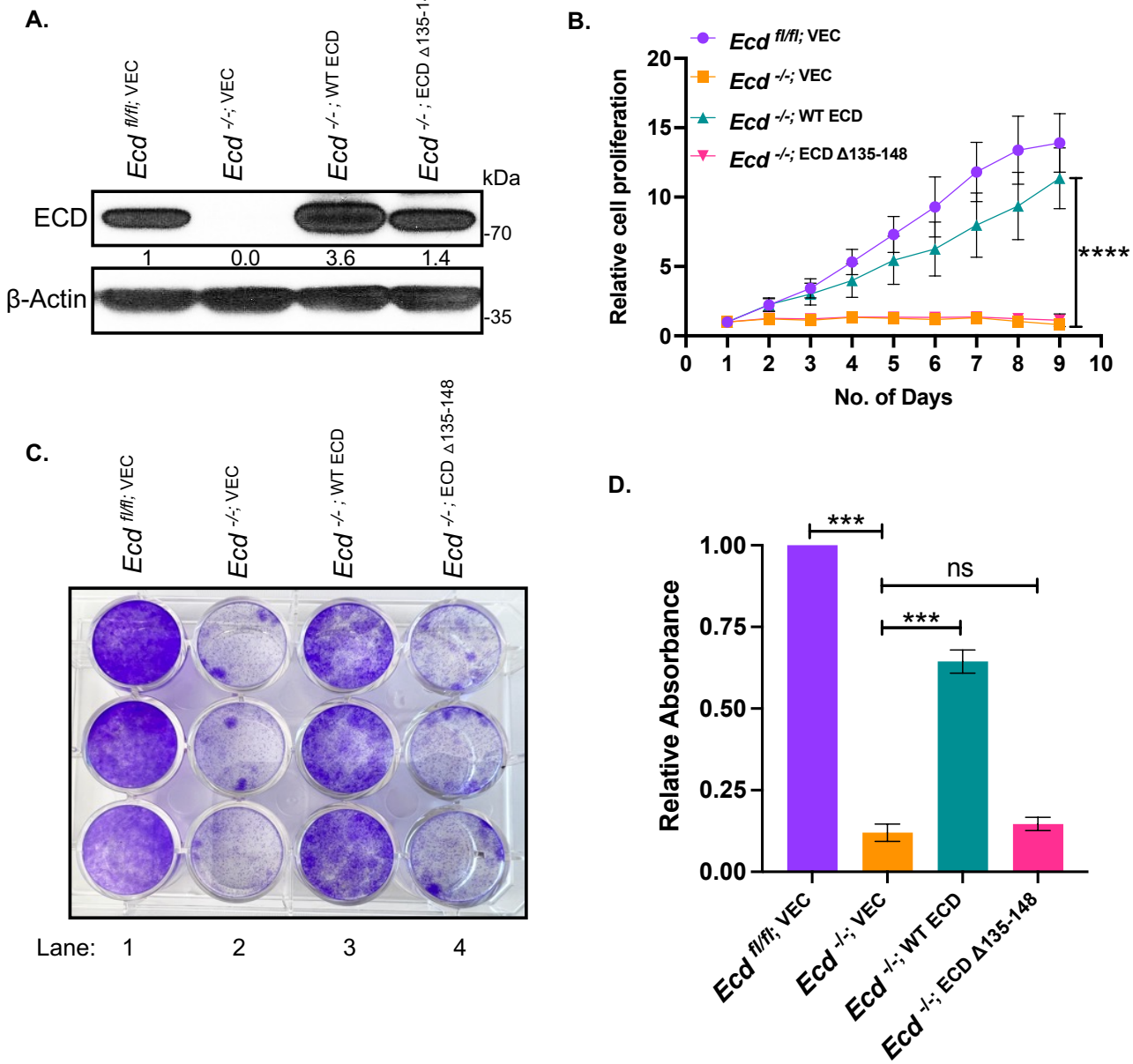
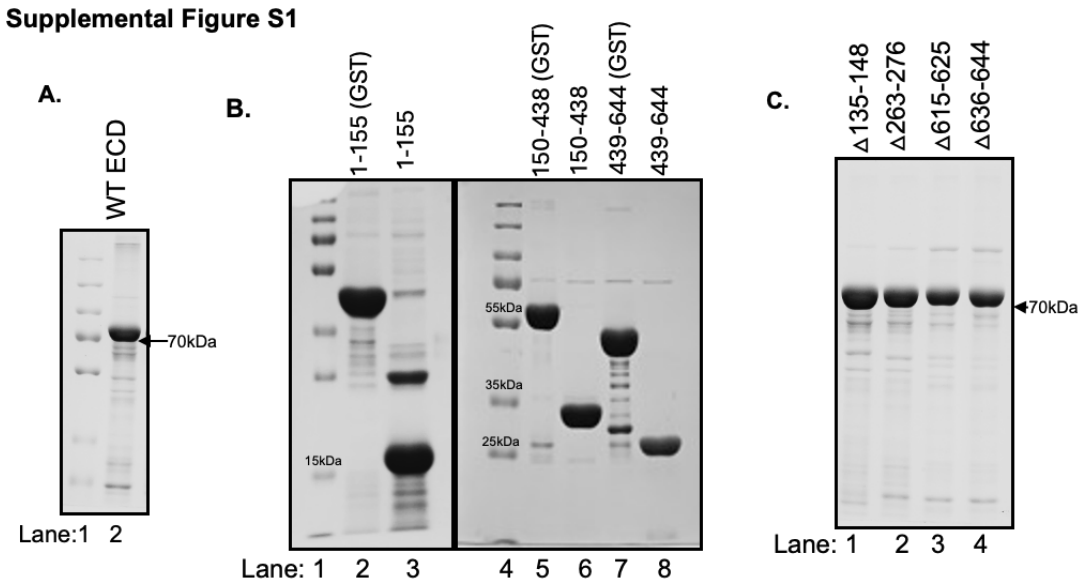

A.
Ecd fl/fl; VEC
Ecd -/-; VEC
Ecd -/- ; WT ECD
Ecd -/- ; ECD Δ135-148
kDa
ECD
-70
1
0.0
3.6
1.4
β-Actin
-35
B.
Ecd fl/fl; VEC
Ecd -/-; VEC
Ecd -/-; WT ECD
Ecd -/-; ECD Δ135-148
Relative cell proliferation
No. of Days
****
C.
Ecd fl/fl; VEC
Ecd -/-; VEC
Ecd -/- ; WT ECD
Ecd -/- ; ECD Δ135-148
Lane: 1 2 3 4
D.
Relative Absorbance
***
ns
***
Ecd fl/fl; VEC
Ecd -/-; VEC
Ecd -/-; WT ECD
Ecd -/-; ECD Δ135-148

Fig. 8. RNA binding defective mutant fails to rescue proliferation block caused by deletion of endogenous ECD. (A) Western blotting shows the expression of WT human ECD or its deletion mutant in *Ecd*$^{fl/fl}$ mouse embryonic fibroblasts (MEFs) with control or Cre adenovirus infection. Numbers below the blot depict band intensities in respect to loading control β-Actin, determined using Image J. **(B)** Proliferation rate of control or ECD deleted MEFs without or with rescue by WT ECD or its RBD mutant (Δ135-148). *Ecd*$^{fl/fl}$ MEFs expressing vector (VEC), WT ECD or its Δ135-148 mutant were infected with vector or Cre adenoviruses to delete endogenous Ecd, followed by cell viability measurement at the indicated time points using CellTiter-Glo based luminescent assay. The data presented as mean values ± SEM from three independent experiments. **** indicate p-value <.0001, computed through two-way Anova. **(C)** Representative images of colony forming abilities of *Ecd*$^{fl/fl}$ MEFs expressing full-length WT ECD or Δ135-148 mutant post infection with vector or Cre adenoviruses, colonies were stained with crystal violet after 10 days, and the solubilized dye absorbance was read at 590 nm. **(D)** Histograms show the relative rescue efficiency of WT ECD and Δ135-148 mutant compared to vector control cells. The data represents mean ± SEM from three independent experiments. *** indicate p-value <.001, ns= non-significant, computed through unpaired student t test.

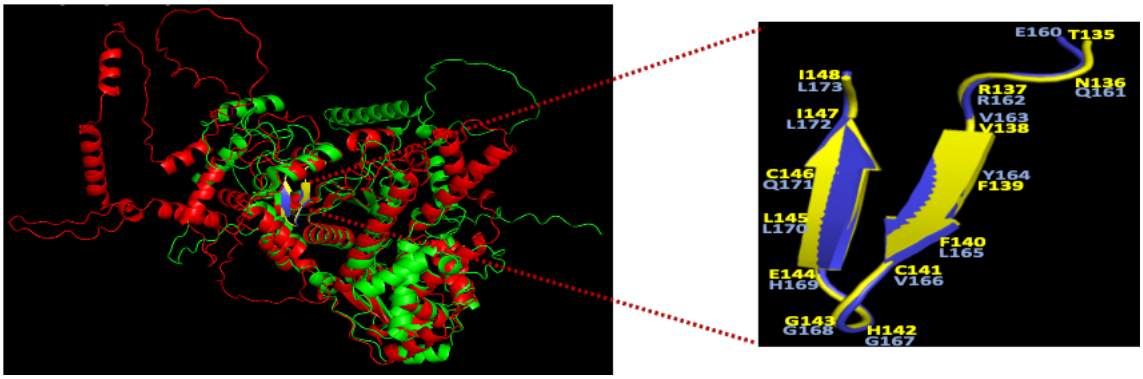
Supplementary Table S1: Nucleotide sequence of Primers and RNA-EMSA Probes.

	Gene	**Forward Sequence (5'-3')**	**Reverse Sequence (5'-3')**
qPCR	*ECD*	ACTTTGAAACACACGAACCTGGCG	TGATGCAGGTGTGTGCTAGTTCCT
	18S	GCTTAATTTGACTCAACACGGGA	AGCTATCAATCTGTCAATCCTGTC
	RNU5A-1	TGGTTTCTCTTCAGATCGCATA	CAAAGCAAGGCCTCAAAAA
Site-directed mutagenesis	ECD∆135-148	CCTGCACCAAGAAAATCTG	GCTATTTTCAGGATCCAG
EMSA Probes	**Probe Name**	**Modification**	**Sequence (5'-3')**
	Probe 1	5'-Biotin	AGCGUGCCGUGCAACAACAUUACAAU UUACAAUCCACCAUGG
	Probe 2	5'-Biotin	UCAACGACUCAACGAC
	U5 snRNA	5'-Biotin	AUACUCUGGUUUCUCUUCAGAUCGCA UAAAUCUUUCGCCUUUUACUAAAGUU UCCGUGGAGAGGAACAACUCUGAGUC UUAACCCAAUUUUUUGAGGCCUUGCU UUGGCAAGGCUA

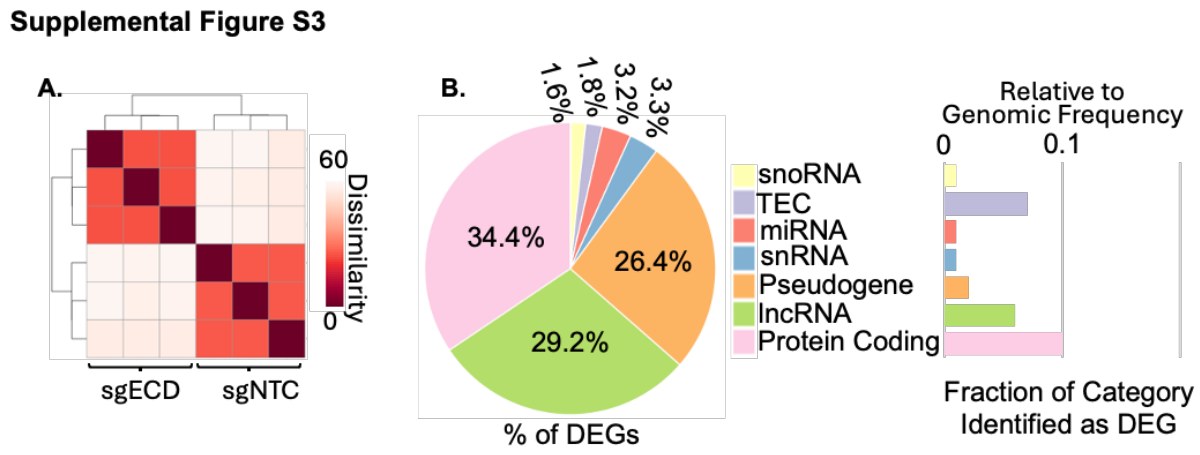


Supplemental Fig. S1. Assessment of purity of recombinant WT ECD and various mutants. Coomassie stained SDS-PAGE gels show the purity of **(A)** wild type (WT) ECD protein. **(B)** truncated mutants of ECD and **(C)** indicated deletion mutants of ECD. All the recombinant proteins expressed in bacteria and purified as GST fusion protein. GST tag was cleaved using PreScission protease.

Supplemental Figure S2

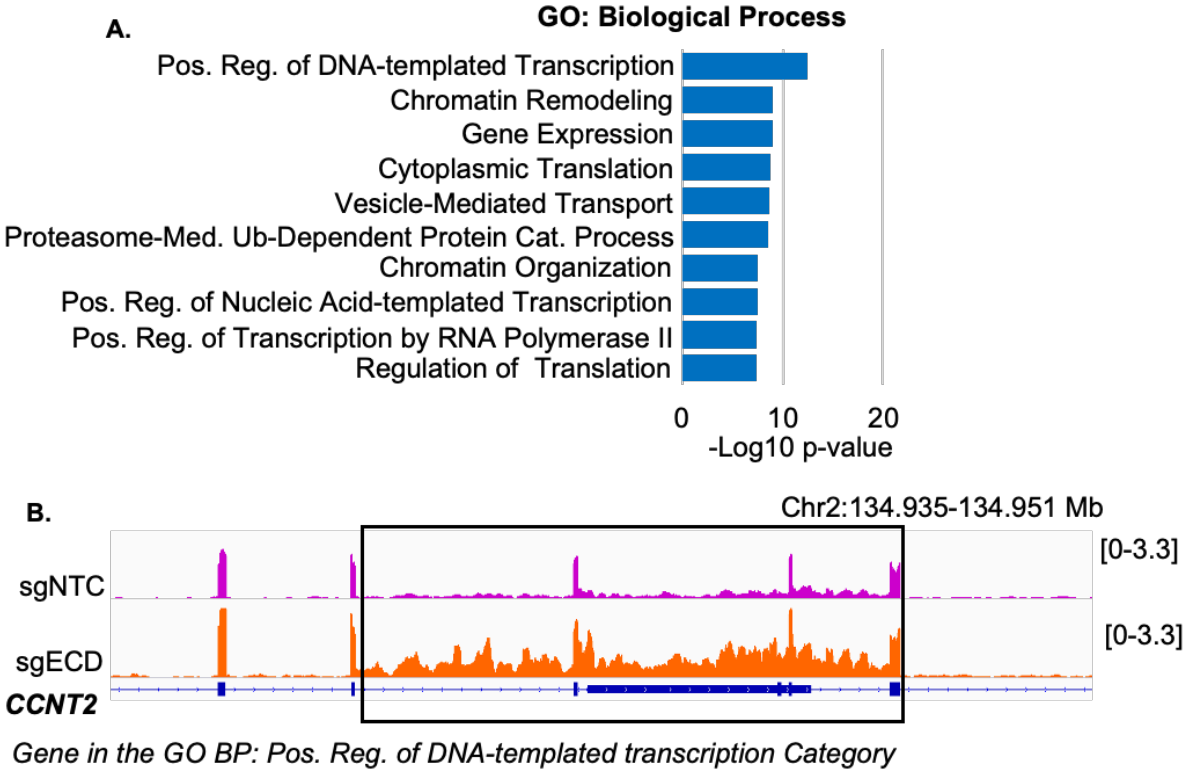


Supplemental Fig. S2. Corresponding amino acids in *Drosophila* Ecd exhibit identical secondary structure to RNA binding region of human ECD. Superimposed predicted tertiary structures of human ECD (green) and *Drosophila* Ecd (red) proteins, showing identified RNA binding region (aa 135-148) of human ECD protein (in yellow) with respective residues (aa 160-173) in *Drosophila* Ecd (in blue). PDB files of Alpha fold predicted 3D structures of human ECD and *Drosophila* Ecd proteins were visualized and superimposed using PyMOL software.



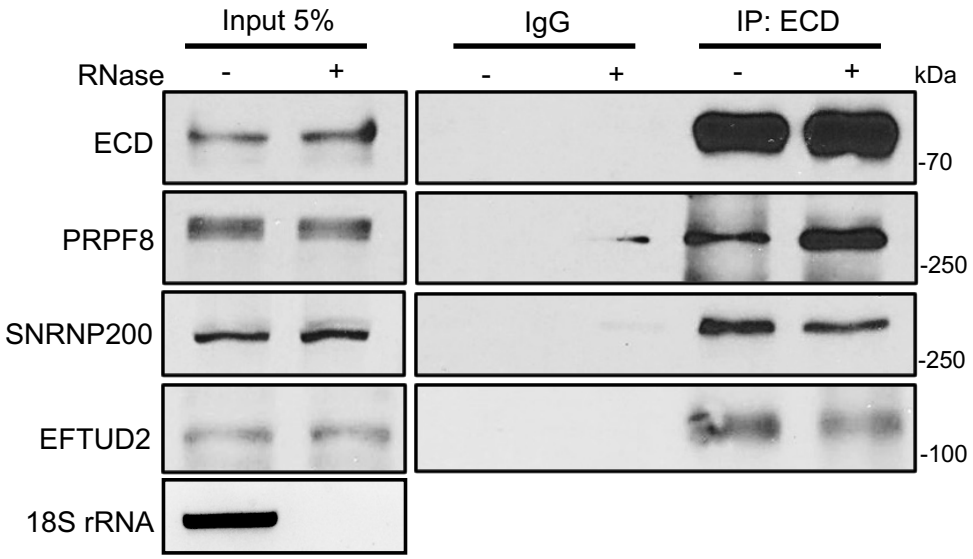
Supplemental Fig. S3. RNA biotypes affected by ECD depletion. A) Replicate comparison of RNA-seq data. **(B)** Categories of annotations for differentially expressed genes showing the raw percentages (left) and the frequency relative to their total presence throughout the genome (right).

Supplemental Figure S4



Supplemental Fig. S4. Pathways affected by differential spliced genes caused by ECD depletion. (A) The top Gene Ontology Biological Processes gene sets associated with differentially spliced genes after ECD depletion. **(B)** RNA-seq signal track showing an aberrant retention of a gene in the top category (shown in A) after ECD depletion.

Supplemental Figure S5



Supplemental Fig. S5. ECD associates to protein components of U5 snRNP complex independent of RNA. Representative Western blots of the indicated proteins after immunoprecipitation using either IgG or ECD specific antibodies in untreated or RNase A treated protein lysates from 76NTERT mammary epithelial cells. Pulldowns through ECD show comparable levels of U5 snRNP protein components in both untreated and RNase A treated conditions. Semi-quantitative PCR of 18S rRNA was used to confirm the efficacy of RNase A treatment.