

# NCBP3 positively impacts mRNA biogenesis

Yuhui Dou<sup>1</sup>, Isabelle Barbosa<sup>2</sup>, Hua Jiang<sup>3</sup>, Claudia Iasillo<sup>1</sup>, Kelly R. Molloy<sup>4</sup>, Wiebke Manuela Schulze<sup>5</sup>, Stephen Cusack<sup>5</sup>, Manfred Schmid<sup>1</sup>, Hervé Le Hir<sup>2</sup>, John LaCava<sup>3,6</sup> and Torben Heick Jensen<sup>1,\*</sup>

<sup>1</sup>Department of Molecular Biology and Genetics, Aarhus University, C.F. Møllers Allé 3, Aarhus 8000, Denmark, <sup>2</sup>Institut de Biologie de l'ENS (IBENS), Département de biologie, École normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France, <sup>3</sup>Laboratory of Cellular and Structural Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA, <sup>4</sup>Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, Rockefeller University, 1230 York Avenue, New York, NY 10065, USA, <sup>5</sup>European Molecular Biology Laboratory, Grenoble Outstation, 71 Avenue des Martyrs, CS 90181, Grenoble Cedex 9 38042, France and <sup>6</sup>European Research Institute for the Biology of Ageing, University Medical Center Groningen, Groningen 9713 AV, Netherlands

Received May 28, 2020; Revised August 19, 2020; Editorial Decision August 20, 2020; Accepted August 26, 2020

## ABSTRACT

The nuclear Cap-Binding Complex (CBC), consisting of Nuclear Cap-Binding Protein 1 (NCBP1) and 2 (NCBP2), associates with the nascent 5' cap of RNA polymerase II transcripts and impacts RNA fate decisions. Recently, the C17orf85 protein, also called NCBP3, was suggested to form an alternative CBC by replacing NCBP2. However, applying protein–protein interaction screening of NCBP1, 2 and 3, we find that the interaction profile of NCBP3 is distinct. Whereas NCBP1 and 2 identify known CBC interactors, NCBP3 primarily interacts with components of the Exon Junction Complex (EJC) and the TRanscription and EXport (TREX) complex. NCBP3–EJC association *in vitro* and *in vivo* requires EJC core integrity and the *in vivo* RNA binding profiles of EJC and NCBP3 overlap. We further show that NCBP3 competes with the RNA degradation factor ZC3H18 for binding CBC-bound transcripts, and that NCBP3 positively impacts the nuclear export of polyadenylated RNAs and the expression of large multi-exonic transcripts. Collectively, our results place NCBP3 with the EJC and TREX complexes in supporting mRNA expression.

## INTRODUCTION

All eukaryotic RNA polymerase II (RNAPII) transcripts undergo processing and protein–RNA packaging events, that are essential for their downstream cellular fates (1). Shortly after initiation of RNAPII transcription, a 7-methylguanosine (m<sup>7</sup>G) cap is enzymatically added to the emerging 5' end of the nascent RNA. While providing protection against 5'–3' exonucleolysis, the cap also serves as

a binding site for the Cap-Binding Complex (CBC), which is composed of the Nuclear Cap-Binding Proteins 1 and 2 (NCBP1/CBP80 and NCBP2/CBP20, respectively) and considered a hallmark of all nuclear RNAPII transcripts (2). NCBP2 directly contacts the 5' cap through its RNA Recognition Motif (RRM), however, high affinity cap-binding is only achieved upon dimerization of NCBP2 and NCBP1 (3,4). Attached to NCBP2, the larger NCBP1 protein serves as a platform for the sequential formation of different complexes engaged with ribonucleoprotein particle (RNP) biology (3,5). In this sense, the CBC impacts various aspects of gene expression, from transcription (6), RNA splicing (2), RNA 3' end processing (7,8), RNA transport (9,10) to RNA decay (11–13), and is hereby capable of influencing RNP fate in both productive and destructive ways (14).

The impact of the CBC on RNP identity is diverse. Early in the RNP assembly process, the CBC interacts with the ARS2/SRRT protein to form the CBC–ARS2 (CBCA) complex, which constitutes a central platform for dynamic protein interactions determining RNA fate (15,16). For example, the CBCA complex may interact with PHAX, which is an essential nuclear transport adaptor for snRNAs and snoRNAs, forming the CBC–ARS2–PHAX (CBCAP) complex (9,12,17–19). If the transcription unit (TU) produces an snRNA, or an independently transcribed snoRNA, transcription termination-coupled RNA 3' end processing will, together with the CBCAP complex, promote the nuclear export or intranuclear transport of the resulting snRNPs and snoRNPs, respectively (9,19). If the TU produces an intron-containing RNA, the CBC will positively impact splicing by interacting with the U4/U6.U5 tri-snRNP and promote spliceosome assembly (2,20). The spliced RNA will then be marked by the Exon Junction Complex (EJC), composed of its

\*To whom correspondence should be addressed. Tel: +45 6020 2705; Fax: +45 8619 6500; Email: thj@mbg.au.dk

four core components EIF4A3, RBM8A/Y14, MAGOH and MLN51/CASC3/BTZ, as well as its multiple peripheral factors (21,22). In addition to the EJC, the Transcription and EXport (TREX) complex is recruited co-transcriptionally (23,24). A central TREX component, ALYREF, can be recruited by the CBC directly to the 5' end of the RNA, bound by the CBCA complex, and joined by other TREX proteins, e.g. DDX39B and the THO complex, together with the EJC, leading to the formation of a nuclear export competent mRNP (10,24–27).

In addition to its effects on RNP production and function, the CBC may also promote RNA nuclear degradation. In the case of such short-lived transcripts, the CBCA complex can link to the zinc-finger protein ZC3H18, which contributes to connect the RNA to either of two exosome adaptors, the Nuclear EXosome Targeting (NEXT) complex (12,28) or the PolyA eXosome Targeting (PAXT) connection (29). NEXT and PAXT both promote RNA degradation via the nuclear RNA exosome and therefore such connections of CBCA-ZC3H18 direct the RNA towards a destructive fate (30). Given this broad repertoire of CBC interactions, it is a central question in nuclear RNA biology how RNP identity is ultimately established. Here, it has been suggested that nuclear RNAs are constantly targeted by factors, often involving mutually exclusive CBC-containing complexes, which ultimately 'settle' RNP fate commitment as a consequence of still ill-defined molecular decision processes (31). Such competing interactions are for example illustrated by the competition between the NEXT/PAXT component MTR4 and ALYREF for association with CBCA-bound RNAs (27) and by the mutually exclusive CBC interactions of ZC3H18 and PHAX (13).

In addition to NCBP1 and NCBP2, the C17orf85/ELG protein was recently shown to bind the m<sup>7</sup>G cap through its putative N-terminal RRM (32). It was further proposed to be able to substitute for NCBP2 in forming an alternative CBC by interacting with NCBP1 via its C-terminal region (32,33); hence it was renamed NCBP3. However, the cap-affinity of NCBP3, unlike that of NCBP2, was not enhanced by NCBP1, and the NCBP1-NCBP3 complex bound the m<sup>7</sup>GTP cap analogue with ~50 fold weaker affinity than that of the canonical CBC (16). Moreover, NCBP3's association with NCBP1 was not mutually exclusive, but rather compatible, with that of NCBP2 (16,32). Consistent with NCBP3 being a cellular partner of ARS2 (16,32), the protein was capable of interacting with the CBCA complex *in vitro* and in a manner preventing formation of the CBC-CAP complex (16). Considering the central concept of factor competition in determining nuclear RNP identity, and to interrogate the alternative CBC hypothesis, we therefore aimed to investigate NCBP3 interactions further.

Applying extensive immunoprecipitation (IP) screening of NCBP1, NCBP2 and NCBP3, we demonstrate a distinct protein interaction network of NCBP3 as compared to its canonical CBC counterparts. Our results identify components of the EJC and TREX complexes as abundant NCBP3 interactors and we show that NCBP3 accumulates in nuclear speckles together with these components. Interestingly, NCBP3 associates with the assembled EJC core and binds RNAs in proximity to the EJC. NCBP3 further aids the TREX complex in mRNA export and positively

regulates the expression of multi-exonic RNAs. We conclude that NCBP3 helps to promote the productive fate of mRNPs, which is supported by its ability to compete with ZC3H18 for RNP association.

## MATERIALS AND METHODS

### Cell lines and cultures

HeLa Kyoto cells stably expressing LAP-tagged proteins (NCBP1, NCBP2, NCBP3 and ALYREF) and control LAP cells were kindly provided by Anthony Hyman (34) (see Supplemental Table S4). NCBP1-, NCBP2-, NCBP3- and control-LAP cells were FACS sorted for EGFP-positive cells. HeLa cells expressing endogenously HA-tagged EIF4A3 were from (35). All cells were maintained in Dulbecco's modified eagle medium (Gibco) supplemented with 10% fetal bovine serum (Sigma) and 100 U/ml penicillin and 100 µg/ml streptomycin (Sigma), at 37°C and 5% CO<sub>2</sub>.

### RNA interference

RNA interference (RNAi) was performed with Lipofectamine 2000 (Invitrogen) according to the manufacturers' instructions. The sequences of the used siRNAs are shown in Supplemental Table S5. Cells were seeded 16 h prior to transfection with Lipofectamine 2000 (10 µl/10 cm dish) and siRNAs (final concentration of 20 nM). Forty eight hours later, the transfection was repeated. Cells were generally harvested 24 h after the second transfection, but for experiments with depletion of EIF4A3, RBM8A, ALYREF and DDX39B, cells were harvested 48 h after the first transfection. For cells used for the RNA-seq analysis, 15 nM of siRNAs was used and cells were harvested 36 h after the second transfection. Depletion efficiencies were monitored by western blotting analysis and protein extractions were performed as described in 'Co-IP experiments' below.

### Co-IP experiments

The interaction screening was performed essentially as described in (36,37). Briefly, three or four replicate IPs were conducted, for each of the extraction conditions. IPs of LAP-tagged NCBP1, NCBP2 and NCBP3 cell lines were carried out in parallel with IP using CTRL-LAP cells. Protein extractions were performed differently for the interaction screen IPs (a) and for the individual IPs (b). (a) Proteins were extracted from cryogrinded cell powders with different extraction solutions supplemented with protease inhibitors (Roche) (solution compositions listed in Supplemental Table S1), at 1:13 (w/v) for 35 mg of NCBP1 and NCBP2, and 1:4 (w/v) for 300 mg of NCBP3, with the aid of brief sonication (QSonica Q700, 1 Amp, 30–40 s; or QSonica S4000, 2 Amp, 15 × 2 s). (b) Cells were lysed in extraction solution (20 mM HEPES, pH 7.4, 0.5% Triton X-100, 150 mM NaCl, supplemented with 1 × protease inhibitors), and sonicated with a microtip sonicator (Branson 250) for 3 × 3 s. After sonication, for both interaction screen IPs and individual IPs, extracts were clarified by centrifugation at 20 000 g and 4°C for 10 min. Supernatants were then incubated with anti-GFP antibodies conjugated to Dynabeads Epoxy

M270 (Invitrogen) (38) and rotated at 4°C for 30 min; or anti-HA antibodies (Abcam, #ab9110) conjugated to protein A Dynabeads (Invitrogen) for EIF4A3-HA IPs and rotated at 4°C for 2 h. Beads were subsequently washed three times with extraction solution and captured proteins were eluted with 1.1 × LDS for SDS-PAGE.

Where RNase treatment was applied, the beads were, after three times of washing, resuspended in 40 µl extraction solution with 1 µl of RNase A/T1 mix (Thermo Scientific # EN0551), and left shaking for 20 min at room temperature, and washed once before elution.

### SDS-PAGE and western blotting analysis

NuPAGE 4-12% Bis-Tris gels (Invitrogen) were run with MOPS or MES buffers (Invitrogen), depending on the sizes of the target proteins, and following the manufacturer's manual. Proteins were subsequently transferred to 0.45 µm PVDF membranes (Immobilon-P membrane) with the iBlot2 transfer system (Thermo Fisher), and blocked with 5% BSA in Tris-buffered saline and 0.1% Tween 20 (TBST). Membranes were then incubated with primary antibody (in TBST with 1% BSA) (antibody list in Supplemental Table S6) at 4°C overnight, followed by washing 3 × 10 min with TBST. HRP-conjugated anti-mouse or rabbit secondary antibodies (Dako), or HRP-conjugated VeriBlot IP Detection Reagent (Abcam) were then applied (1:10 000 in TBST with 1% BSA), and incubated at room temperature for 1 h. After washing as above, membranes were incubated with ECL substrate (Thermo Scientific #34096) and imaged with Amersham Imager 600.

### MS and data processing

Details of MS sample processing is described elsewhere (37). Briefly, samples were reduced, alkylated and in-gel trypsin digested. Peptides were then extracted with 1% TFA and cleaned with C18 tips, after which they were analysed with liquid chromatography and MS on Orbitrap Fusion or Q Exactive instruments (Thermo Fisher Scientific). MS raw data were processed with the MaxQuant package, with 'label-free quantification (LFQ)' and 'match between runs' enabled (39). The Maxquant output was processed with Perseus software, where proteins labelled as 'only identified by site', 'reverse' or 'contaminant' were filtered out. LFQ intensities were log<sub>2</sub> transformed and proteins were filtered with valid values in at least two out of all of the replicates. The missing values were imputed from normal distributions with default Perseus setting. *t* tests were then performed between each NCBP IP and the control IP. Proteins with a false discovery rate (FDR) < 0.01, *S*<sub>0</sub> (log<sub>2</sub> fold changes) > 1 were considered as significantly enriched. Using significantly enriched proteins, stoichiometric abundances were calculated as follows (40): (i) mean LFQ intensity of the replicates was divided by the molecular weight of the protein; (ii) the background of the control IP was subtracted; (iii) the resulting value was normalized to that of the bait, which was set to 100. Hierarchical clustering was performed using log<sub>2</sub>(stoichiometric abundance × 10<sup>3</sup>) with Euclidean distances under Perseus default settings. The Venn diagram was generated with the BioVenn tool (41).

### Recombinant protein binding assay

The purified recombinant proteins and the performance of interaction assays were previously described (16,42). Briefly, 3 µg of each protein was mixed in Binding Buffer BB-125 in a final volume of 60 µl (125 mM NaCl final concentration) complemented, or not, with 2 mM ADPNP and/or single-stranded biotinylated RNA (10 µM). After incubation for 20 min at 30°C, 12 µl of pre-blocked m<sup>7</sup>G affinity beads (50% slurry, 7-Methyl-GTP Sepharose<sup>®</sup> 4B, GE Healthcare) and 200 µl of BB-250 (250 mM NaCl) were added. After gentle rotation for 2 h at 4°C, the beads were washed three times with BB-150 (150 mM NaCl) and eluted with 1 × SDS loading dye. Eluates were boiled and loaded on 4-12% SDS-PAGE gels with a protein marker (PageRuler unstained Protein Ladder, Fermentas). Proteins were visualized by Coomassie staining.

### Immunofluorescence and microscopy

For protein immunofluorescence, cells were grown on 18 mm cover slips, washed twice with PBS and fixed with 4% paraformaldehyde in PBS for 12 min. They were then washed twice with PBS, permeabilized and blocked with 0.5% Triton X-100 and 3% FBS in PBS for 15 min and washed three times with PBS. Cells were then incubated with primary antibodies for 1 h and washed three times with PBS before incubation with secondary antibodies for 45 min followed by three times of washing with PBS (antibody list shown in Supplemental Table S7). Finally, cells were stained with DAPI 1 µg/ml for 10 min, followed by washing with PBS twice and H<sub>2</sub>O once before mounting on slides with ProLong Gold antifade reagent. pA<sup>+</sup> RNA fluorescent *in situ* hybridization (FISH) analysis was performed as previously described (43) using a Cy5-labeled oligo (dT)<sub>70</sub> probe (DNA Technology A/S). Images were obtained with a ZEISS Axio Observer 7 microscope equipped with an AxioCam 702 camera and Plan-APOCHROMAT 63 ×/1.4 oil objective, and Zen 2 software.

### RNA extraction and RT-qPCR analysis

For RNA used for RT-qPCR analysis, extraction was performed using TRIzol (Invitrogen) following the manufacturer's instructions. RNA was DNase treated with TURBO DNA-free kit (Invitrogen) before reverse transcription with SuperScript III Reverse Transcriptase (Invitrogen). The cDNA obtained was used for qPCR with SYBR Green qPCR SuperMix (Invitrogen) and AriaMx Real-time qPCR System (Agilent). Sequence of primers used is listed in Supplemental Table S8.

### RNA sequencing (RNA-seq) and data analysis

RNA extraction and RNA-seq library preparation upon GFP and NCBP3 depletion were performed as described (44). Total RNAseq of siNCBP3 depletion samples are first reported here, but were collected as part of the same experiment described in (44,45). Reads from siNCBP3 samples were processed in parallel with the siEGFP control and relevant depletions from GEO:GSE99059 were processed in parallel as described in (46). In brief,



raw reads were quality filtered and trimmed as described (29), using Trimmomatic (v 0.32) and settings (PE ILLUMINACLIP:/com/extra/Trimmomatic/0.32/adapters/TruSeq3-PE-2.fa:2:30:10 HEADCROP:12 LEADING:22 SLIDINGWINDOW:4:22 MINLEN:25). Cleaned reads were then mapped to GRCh38 with HISAT2 (v 2.1.0) (47) using default settings and the genome index 'H. sapiens, UCSC hg38 and Refseq gene annotations' provided at the HISAT2 download page ([ftp://ftp.ccb.jhu.edu/pub/infphilo/hisat2/data/hg38\\_tran.tar.gz](ftp://ftp.ccb.jhu.edu/pub/infphilo/hisat2/data/hg38_tran.tar.gz)). Only proper pairs with both reads mapping to the genome were used for further analysis. Exon read counts were collected using the featureCounts tools from the subread software suite (v 2.0.0) (48) with settings '-p -C -s 2 -t exon' for the default GRCh38 annotations for featureCounts. Differentially expressed transcripts were obtained from raw read counts using R package DESeq2 (v 1.20.0) at a false discovery rate (FDR) cutoff of 0.1. Exon counts for differentially up- and down-regulated transcripts were compared using custom scripts in R.

### CLIP data analysis

Processed CLIP data for NCBP3 and EIF4A3 were obtained from (49). Data shown are from NCBP3 PAR-CLIP data processed with the PARalyzer tool and the EIF4A3 HITS-CLIP data was processed with Piranha (49). NCBP2 CLIP data was from GSE94427 (13) and lifted to GRCh38. ALYREF CLIP data was from GSE113896 (24) and lifted to GRCh38. Metagene plots for CLIP data were constructed using custom python and R scripts employing deeptools software (v3.0.2) (50) relative to GRCh38 annotations shipped with featureCounts used for RNAseq analysis. For T>C conversion positions, raw reads were obtained from SRA: SRR500480 and SRR500481 (51), adapter sequences (TCTTTTATCGTATGCCGCTCTTCTGCTTG and TCTCCATCGTATGCCGTCTTCTGCTTG), low quality and short reads removed using bbduk from the BBMAP software (v35.92, <https://jgi.doe.gov/data-and-tools/bbtools/>) using parameters  $k=13$ ;  $ktrim=r$ ;  $useshortkmers=t$ ;  $mink=5$ ;  $qtrim=t$ ;  $trimq=10$ ;  $minlength=20$ ;  $ref=barcodes.fa$ , where the 'barcodes.fa' file contains the 2 adapter sequences. Remaining reads were mapped to GRCh38 using the STAR aligner (v2.5.2b) (52) together with samtools (v1.6) (53) with settings -outFilterType BySJout; -outFilterMultimapNmax 20. Duplicate reads were removed using samtools rmdup, remaining mismatched bases collected using samtools mpileup and T to C conversion positions extracted from mpileup output using a custom script counting T>C mismatches for plus strand conversions and A>G mismatches for minus strand conversions.

## RESULTS

### Interrogating interactions facilitated by NCBP1, NCBP2 and NCBP3

If NCBP1 and NCBP3 would form an alternative CBC (32), we surmised that their interaction would be independent of, or perhaps even competitive with, NCBP2. We therefore conducted co-IP experiments of NCBP1, employing

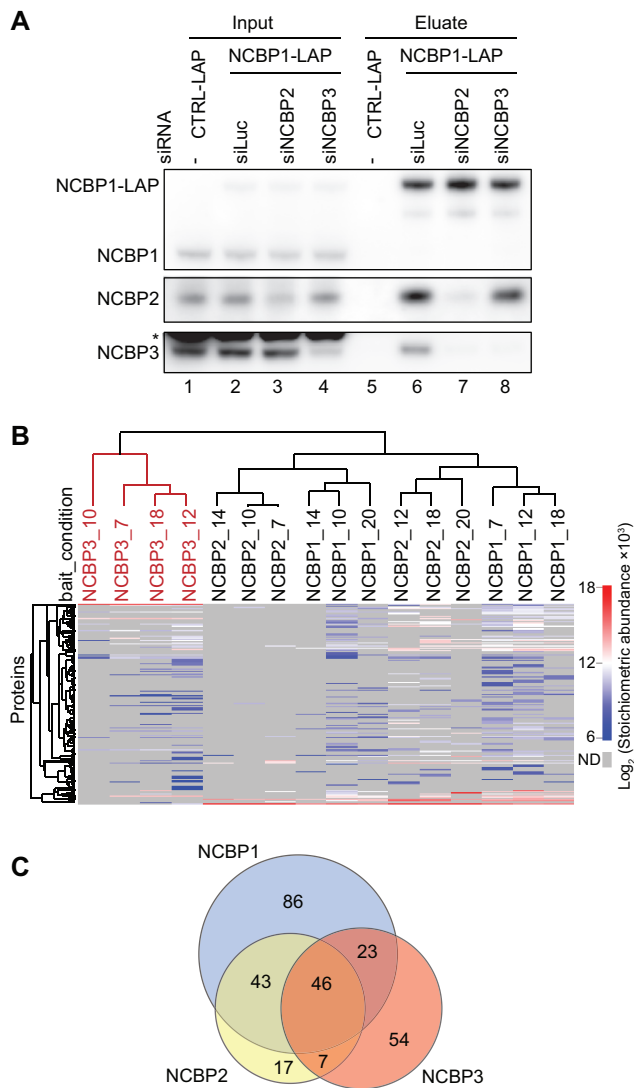
HeLa cells that express Localization and Affinity Purification (LAP)-tagged NCBP1, upon siRNA-mediated individual depletion of NCBP2 or NCBP3. The NCBP1-NCBP2 interaction was virtually lost upon NCBP2 depletion (Figure 1A, compare lanes 6 and 7). Moreover, the NCBP1-NCBP2 interaction was unaffected by NCBP3 depletion. Thus, despite *in vitro* studies showing that NCBP3 can bind NCBP1 in the absence of NCBP2 (16,32), it appears that the NCBP1-NCBP3 interaction *in vivo* requires the presence of NCBP2.

As we wanted to further explore the interaction networks of NCBP1, NCBP2 and NCBP3, we next conducted IP-mass spectrometry (IP-MS) analysis using HeLa cells stably expressing LAP-tagged versions of the three NCBP3s at near endogenous level (Supplemental Figure S1), or the LAP tag alone as a negative control (34). To obtain comprehensive interaction profiles, we applied an optimized IP screen, where multiple extraction conditions were used to favor a broad range of protein interactions (36). Details of this IP screen are described elsewhere (37). After a preliminary screen with 24 different extraction conditions, final IPs were conducted in triplicate or quadruplicate using six different conditions for NCBP1 and NCBP2, and four different conditions for NCBP3 and the control IP (see Supplemental Table S1 for solution details). Proteins that were enriched with statistical significance in NCBP-LAP IPs over the LAP control IPs, were identified (54). Their stoichiometric abundances, relative to the relevant bait protein, were then calculated (40) and used to perform hierarchical clustering over the different conditions. Remarkably, the NCBP3 interaction profile formed a distinct cluster, whereas NCBP1 and NCBP2 interactions clustered together (Figure 1B). Consistently, ~80% of proteins captured by NCBP2 were common with NCBP1, whereas NCBP3 only shared ~50% of its interactors with NCBP1, with the majority of the rest being NCBP3-specific (Figure 1C).

Taken together, our results demonstrate that NCBP3 harbors a partially distinct interaction profile and suggest that the protein only associates with NCBP1 and NCBP2 in their CBC context.

### NCBP3 predominantly interacts with EJC and TREX components

Further analyses of the stoichiometric interaction profiles of NCBP proteins revealed that NCBP1 and NCBP2 primarily interacted with each other and with PHAX (Supplemental Figure S2). In contrast, the most predominant interactors of NCBP3 were components of the EJC and TREX complexes, among which EJC core components EIF4A3, RBM8A and MAGOH were most abundantly enriched across all IP conditions (Supplemental Figure S2). To better compare interaction differences between each of the NCBP proteins, we summarized protein groups/complexes that abundantly interact with NCBP3 or NCBP1 and NCBP2 in a heatmap (Figure 2A, for extensive complex enrichment comparison (37)). PHAX was exclusively, and ARS2 was primarily, enriched in NCBP1 and NCBP2, but not in NCBP3, IPs, which was consistent with previous studies (16,32). Another group of proteins that stood out as being enriched in NCBP1 and NCBP2, but not in NCBP3, IPs,



**Figure 1.** Protein-protein interactions facilitated by NCBP1, NCBP2 and NCBP3. (A) Western blotting analysis of NCBP1-LAP co-IPs from HeLa cell extracts upon depletion of luciferase (siLuc) control, NCBP2 or NCBP3 using the indicated siRNAs. IPs were performed with anti-GFP antibodies. A negative control IP ('CTRL-LAP') was conducted using cells expressing the untagged LAP moiety. Lanes 1–4 and lanes 5–8 show input and eluate samples, respectively. Asterisk indicates unspecific protein band. Gel loading: 0.05% of input, 15% of eluate. (B) Hierarchical clustering (Euclidean distances) of log<sub>2</sub> transformed stoichiometric abundances derived from NCBP1, NCBP2 and NCBP3 interaction screening over various extraction conditions (see Supplemental Table S1). Columns represent the clustering of individual NCBP IPs (with each extraction condition signified by a number) with the NCBP3 cluster highlighted in red. Rows indicate the protein interactors identified. To calculate stoichiometric abundance, mean Label Free Quantification (LFQ) intensity was normalized to the molecular weight of the protein and deducted that of the negative control sample, the value of which was then further normalized to that of the relevant bait protein. Only proteins passing a T-test significance threshold of (FDR < 0.01 and log<sub>2</sub> fold change > 1) were considered. (C) Venn diagram displaying the overlap of proteins captured by NCBP1-LAP (blue), NCBP2-LAP (yellow) and NCBP3-LAP (red) that were significantly enriched in at least one of the six extraction conditions.

were factors related to RNA decay by the nuclear exosome, including ZC3H18, MTR4/SKIV2L2 and ZCCHC8. The heatmap also displayed that EJC and TREX components are abundant NCBP3 interactors exhibiting substantially enhanced co-enrichment with NCBP3 compared to NCBP1 and 2.

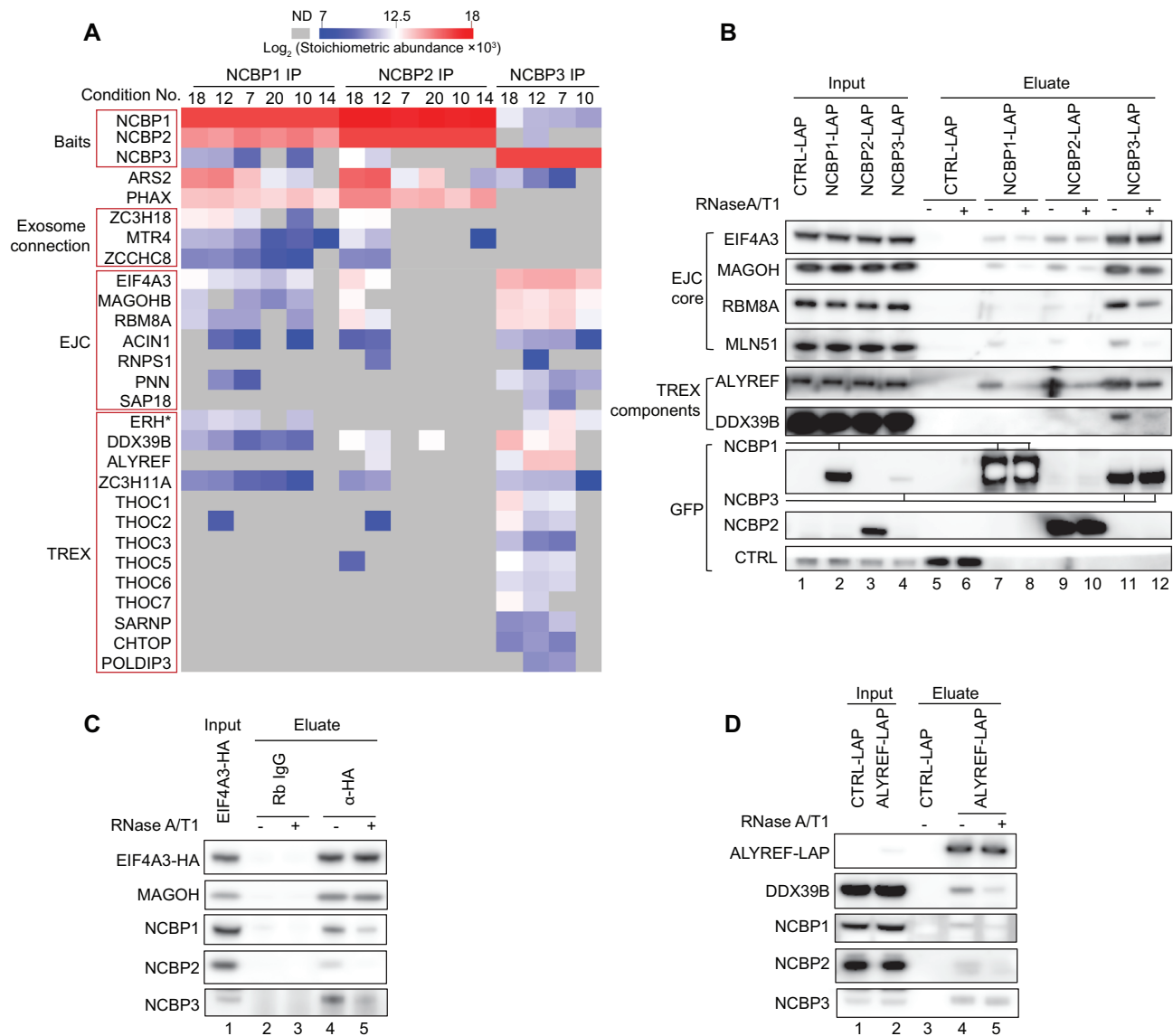
We next sought to confirm the interactions of NCBP3 with EJC and TREX components by IP/western blotting analysis. While all three NCBP proteins were able to co-IP EIF4A3, RBM8A and MAGOH (EJC core components) as well as ALYREF and DDX39B (TREX components), NCBP3 did this most efficiently; this was even more striking considering that NCBP3-LAP expression and capture levels were discernibly lower than those of NCBP1-LAP and NCBP2-LAP (Figure 2B). It was also notable, that the EJC core component MLN51 was not identified in the NCBP3 IP-MS screening, yet it appeared in the IP-western validation (Figure 2B, lanes 7, 9 and 11). However, compared to the three other EJC core components, MLN51 was captured less efficiently. This finding might be explained by the largely cytoplasmic localization of MLN51, whereas the remaining EJC core components and NCBP3 mainly reside in the nucleus (32,55,56). It is notable that NCBP3's interaction with the EJC components EIF4A3 and MAGOH was not sensitive to RNase treatment, whereas its interaction with RBM8A and MLN51 was. Finally, we also performed reverse-IP/western analysis using HeLa cells stably expressing HA-tagged EIF4A3 or LAP-tagged ALYREF. Both EIF4A3-HA and ALYREF-LAP IPs recapitulated previous data by co-purifying NCBP3 in an only mildly RNase-sensitive manner (Figure 2C, D).

We conclude that the most abundant interactors of NCBP3 are EJC and TREX components. Unlike NCBP1 and NCBP2, NCBP3 does not copurify PHAX and exosome adaptors, suggesting that it neither functions in sn(o)RNA transport nor in exosomal RNA decay.

### EJC core integrity is required for its interaction with NCBP3

To further characterize the interaction dependence of the EJC and TREX complexes with NCBP3, we conducted NCBP3-LAP IPs upon depletion of selected EJC and TREX components. As previously reported, only partial depletion was achieved for the EJC components EIF4A3 and RBM8A (57), presumably due to the essential nature of these factors (Figure 3A, lanes 1–3). Still, their decreased expression resulted in reduced IP levels of the remaining two EJC core factors EIF4A3 and MAGOH/RBM8A (Figure 3A, lanes 6–8). This suggests that trimeric EJC core integrity is important for NCBP3 interaction. MLN51 IP levels were unaffected by EIF4A3 or RBM8A depletion. For the TREX components ALYREF and DDX39B, their partial depletions did not lead to a reduction of the captured EJC core proteins (Figure 3A, lanes 4–5 and 9–10).

We next interrogated the EJC core integrity-dependent NCBP3 interaction further, by conducting *in vitro* protein binding assays of recombinant EJC core proteins and NCBP3. Taking advantage of the affinity of NCBP3 to m<sup>7</sup>-GTP, we used m<sup>7</sup>-GTP Sepharose to purify NCBP3-bound proteins (16,32). This revealed that NCBP3 was unable to associate with either of the individual EJC core proteins



**Figure 2.** NCBP3 predominantly interacts with EJC and TREX components. (A) Heatmap comparing abundances of selected protein groups among the three NCBP protein IPs as indicated. The heatmap shows log<sub>2</sub> transformed stoichiometric abundances of proteins significantly enriched in the NCBP IPs across different extraction conditions. Proteins known to belong to certain complexes are grouped in labelled boxes on the left. The heat color range is shown on the top and missing values (ND) are shown in grey. \*ERH as a putative TREX components (23,58). The full data set of significantly enriched interactions of NCBP proteins with the indicated EJC and TREX components. IPs were performed on extracts from control (CTRL-LAP), NCBP1-LAP, NCBP2-LAP and NCBP3-LAP HeLa cells with (+) or without (-) RNase A/T1 treatment. Bait proteins were detected with anti-GFP antibodies. Lanes 1-4 and 5-12 show input and eluate samples, respectively. Gel loading: 0.05% of input, 15% of eluate. (C) Western blotting analysis of the indicated proteins after EIF4A3-HA co-IP with HA antibody (Rabbit IgG as negative control). RNase A/T1 treatments were as indicated. Gel loading: 0.05% of input, 15% of eluate. (D) Western blotting analysis of the indicated proteins after ALYREF-LAP co-IP with anti-GFP antibody (CTRL-LAP cell extract as negative control). RNase A/T1 treatments were indicated. Gel loading: 0.05% of input, 15% of eluate.

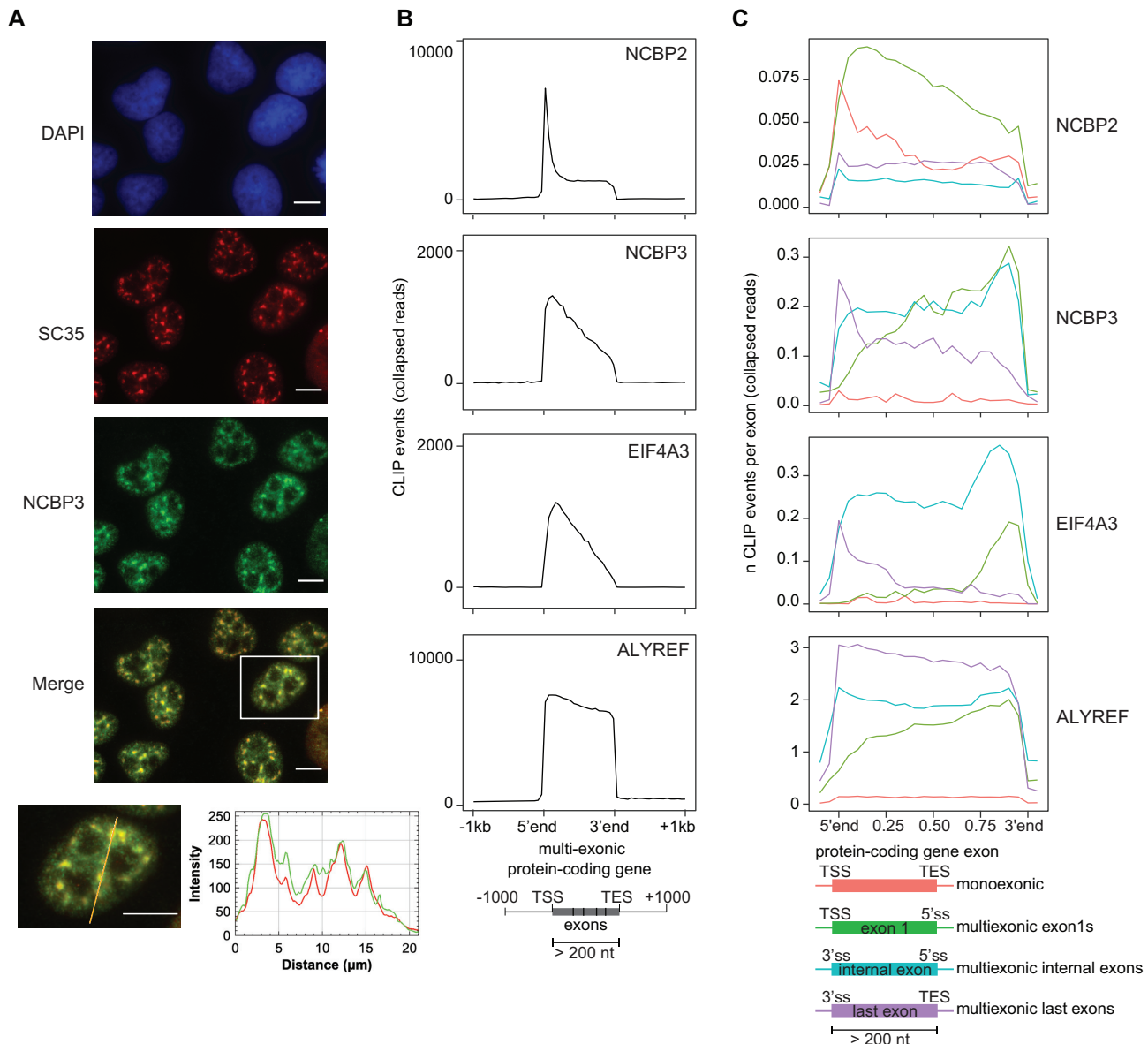
(MAGOH/RBM8A added as a heterodimer) (Figure 3B, lanes 5–7), or with a simple mix of these individual core components (Figure 3B, lane 8). Only when the EJC core, with all four of its components, was reconstituted on RNA and locked in its closed conformation upon addition of the non-hydrolysable ATP analogue ADPNP (42), was NCBP3 able to bind (Figure 3B, lane 9). Hence, although MLN51 interaction with NCBP3 was less abundant compared to the other EJC core components (Figure 2B), and appeared insensitive to EIF4A3 or RBM8A depletion (Figure 3A), the protein was still required for EJC interaction with

NCBP3 *in vitro* (Figure 3B, compare lanes 9 and 10) (see Discussion). We also performed pull down assays with an NCBP3 construct comprising only the aa1–128 N-terminal region, which harbors the RRM of the protein, conferring cap and ARS2 affinity (16,32). Although slightly less efficient than the full-length protein, this N-terminal fragment was also able to pull down the EJC core (Supplemental Figure S3).

Taken our data together, we conclude that NCBP3 binds to the EJC, but only when it is fully assembled, which suggests that the interaction occurs post-splicing.







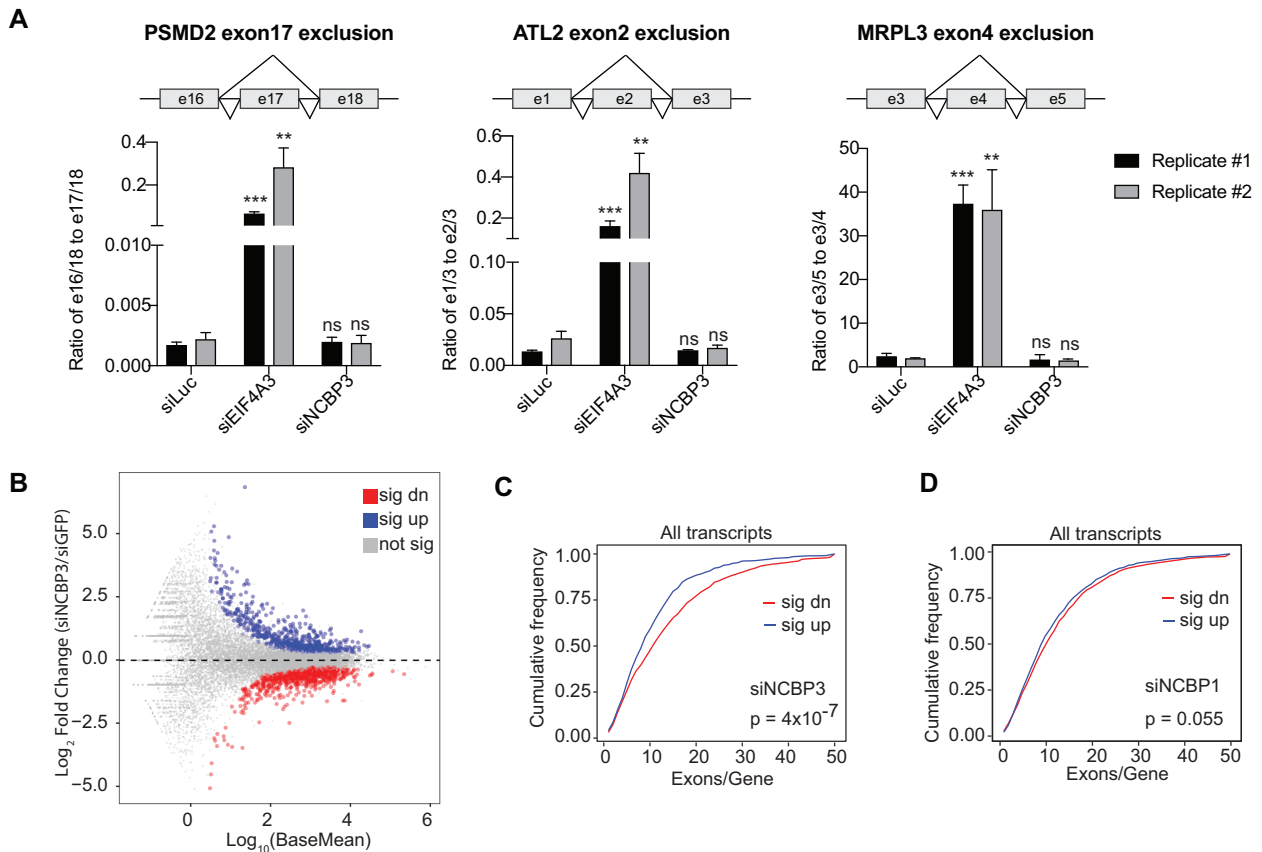
**Figure 4.** NCBP3 co-localizes with EJC and TREX complexes *in vivo*. (A) Immunolocalization microscopy of NCBP3 and the nuclear speckle marker SC35 in HeLa cells, using antibodies. DAPI, SC35 and NCBP3 stainings are shown in blue, red and green, respectively. SC35 and NCBP3 signals were merged and a zoomed-in cell is shown with a line scan profile along the drawn yellow line. Scale bars are 10  $\mu\text{m}$ . (B) CLIP signal profiles of NCBP2, NCBP3, EIF4A3 and ALYREF over exons of protein coding transcripts from their transcription start sites (TSSs) to transcript end sites (TESs). (C) CLIP signal profiles as in (B) but over the exons (using a subset of exons >200nts long) of monoexonic (red) as well as first (green), internal (blue) or last (purple) exons of multiexonic protein coding transcripts.

### NCBP3 depletion causes exon-number biased regulation of RNA expression

The relationship between NCBP3 and EJC/TREX prompted the question whether NCBP3 might function in the context of EJC and/or TREX. To probe a possible functional relation of NCBP3 to the EJC, we took advantage of the circumstance that depletion of EJC core components, or the peripheral component RNPS1, affects certain alternative splicing decisions (57,65). However, examining the effect of NCBP3 depletion on three exon exclusion events, that were all sensitive to EIF4A3 depletion, yielded no discernible phenotypes (Figure 5A).

We then asked whether NCBP3 impacts global RNA expression levels by sequencing total RNA (RNA-seq) from HeLa cells depleted of NCBP3 (Supplemental Figure S5A and Supplemental Table S3). Differential expression analysis identified 724 and 818 annotated transcripts that were significantly up- or down-regulated, respectively, upon NCBP3 depletion (Figure 5B). We noticed that transcripts containing multiple exons (>10) were biased towards being downregulated (Figure 5C), and that this trend was dominated by protein coding transcripts (Supplemental Figure S5B). Employing RNA-seq data collected in parallel (44,45) demonstrated that NCBP1 depletion did not show such an





**Figure 5.** NCBP3 positively impacts the expression of multi-exonic transcripts. (A) RT-qPCR analysis of the schematized exon skipping events upon depletion of Luciferase (siLuc), EIF4A3 (siEIF4A3) or NCBP3 (siNCBP3) as indicated, and using primers listed in Supplemental Table S8. Y-axes show the ratios of exon skipping isoforms relative to non-skipping isoforms. Biological duplicate experiments are shown, with data presented as mean values and error bars indicating standard deviation ( $n = 3$  technical qPCR replicates).  $P$  values were obtained from unpaired  $t$ -tests, two-tailed, with  $**P \leq 0.01$ ,  $***P \leq 0.001$ , and ns as 'not significant'. (B) MA-plot showing the population of significantly up- (blue) and down-regulated (red) annotated transcripts from HeLa cells subjected to NCBP3- relative to eGFP (control)-depletion. Log<sub>2</sub> fold changes on the Y-axis are plotted against the log<sub>10</sub> mean of normalized counts of control and siNCBP3 samples ('baseMean', X-axis). Unaffected RNAs are shown in grey. (C) Cumulative frequency plot of exon numbers in upregulated (blue) and downregulated (red) transcripts upon NCBP3 depletion. The  $P$  value from a Mann-Whitney U-test comparing exon numbers in significantly up- versus down-regulated transcripts is indicated. (D) Cumulative frequency as in (C) but for NCBP1 depletion.

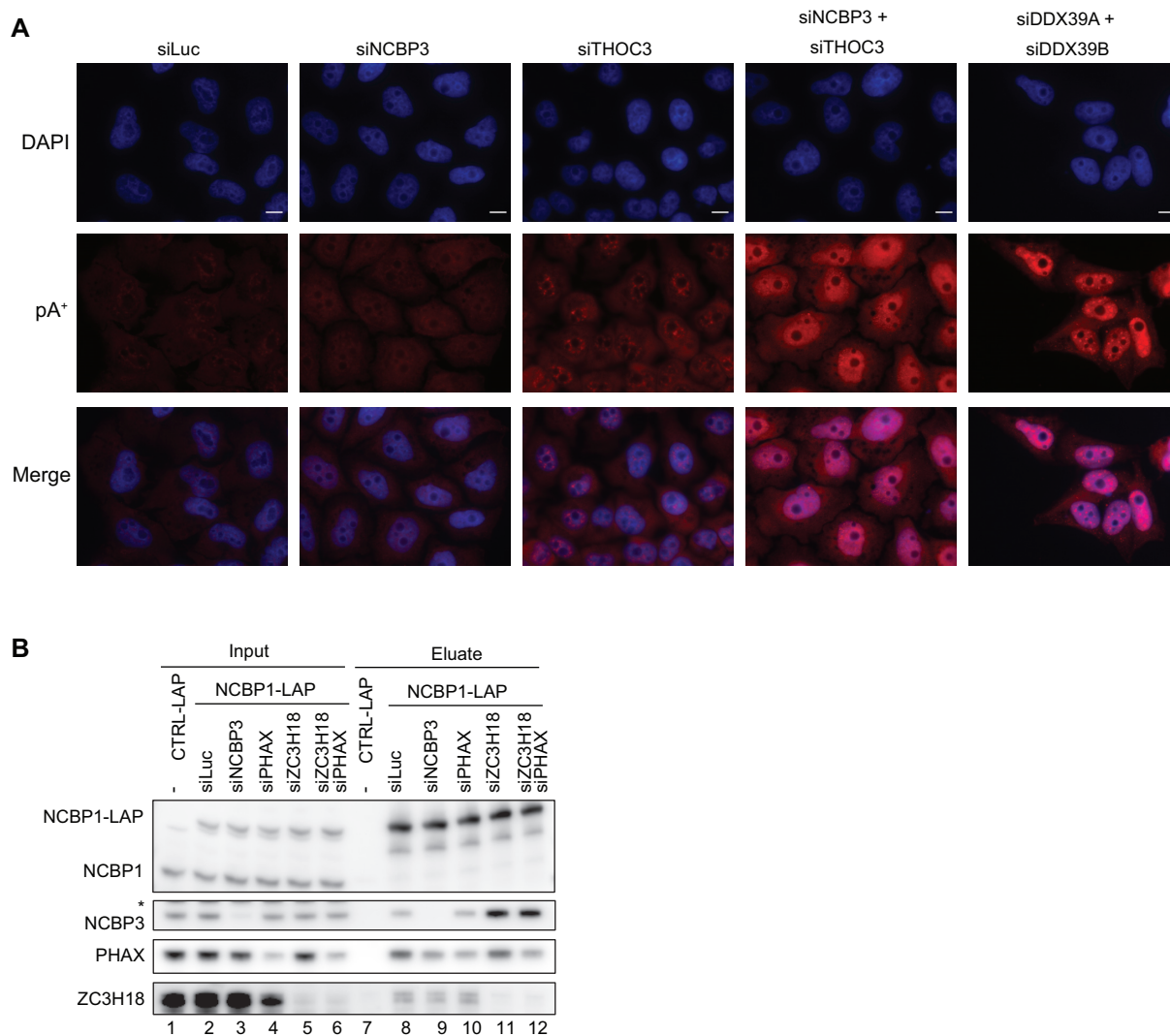
exon-content effect on RNA levels (Figure 5D, Supplemental Figure S5B). A possible caveat to this finding was that expression levels of upregulated transcripts were significantly lower than for those that were downregulated (Supplemental Figure S5C), and that RNAs containing many exons generally were present in higher levels than those with fewer exons (Supplemental Figure S5D). However, using a set of expression-matched up- and down-regulated transcripts revealed a similar trend, supporting that the phenotype of NCBP3 depletion was exon- but not expression-dependent (Supplemental Figure S5E).

We conclude that NCBP3 does not partake in core EJC functions, yet the protein positively impacts the expression of multi-exonic transcripts. Whether this exon number-dependent effect relates to transcription, RNA decay or RNA localization remains unclear.

### NCBP3 contributes to the nuclear export of polyadenylated RNA

We then proceeded to inquire about a functional connection of NCBP3 with the TREX complex. Since individual

depletion of NCBP3 does not elicit a prominent polyadenylated (pA<sup>+</sup>) RNA export defect (32), we sought for a possible synthetic effect between NCBP3 and TREX inactivation. Depletion of individual TREX components often results in strong nuclear accumulation of pA<sup>+</sup> RNA (66), so we selected the THOC3 subunit whose depletion has a relatively mild phenotype (66), and performed fluorescent *in situ* hybridization (FISH) for pA<sup>+</sup> RNA upon single- and double-depletion of NCBP3 and THOC3 (Figure 6A and Supplemental Figure S6). Co-depletion of the TREX components DDX39A and DDX39B was conducted as a positive control (67). As demonstrated in previous studies (32,66), individual depletion of NCBP3 and THOC3 did not yield strong nuclear pA<sup>+</sup> RNA accumulation (Figure 6A). Instead, co-depletion of NCBP3 and THOC3 yielded a strong effect. Although the possibility of nuclear RNA stabilization and/or lengthening of pA tails exist, we find it likely that NCBP3 plays a role in RNA export in conjunction with TREX. It is also notable that the nuclear retained pA<sup>+</sup> RNA, upon co-depletion of NCBP3 and THOC3, did not show as strong a nuclear speckle accumulation as upon DDX39A/DDX39B co-depletion. Whether this reflects an



**Figure 6.** NCBP3 positively impacts pA<sup>+</sup> RNP export competence. (A) Fluorescence *in situ* hybridization (FISH) analysis of pA<sup>+</sup> RNA in HeLa cells after individual- or co-depletion of factors with the indicated siRNAs. Depletions were performed for 72 h with siLuc (control), siNCBP3, siTHOC3, and co-depletion with siNCBP3 and siTHOC3. For siDDX39A and siDDX39B depletion was performed for 48 h. DAPI and pA<sup>+</sup> RNA signals are shown in blue and red, respectively. Merge is shown at the bottom. Scale bars are 10  $\mu$ m. (B) Western blotting analysis of NCBP1-LAP co-IPs from HeLa cell extracts depleted of luciferase (control), NCBP3, PHAX, ZC3H18 or both PHAX and ZC3H18 as indicated. Lanes 1–6 and lanes 7–12 show input and eluate samples, respectively. Asterisk indicates unspecific protein band. Gel loading: 0.05% of input, 15% of eluate. IPs were performed using an anti-GFP antibody.

impact of THOC3/NCBP3 on pA<sup>+</sup> RNA progress outside of nuclear speckles remains an interesting question to be investigated.

A possible function of NCBP3 in the export of pA<sup>+</sup> RNA, taken together with the absence of exosome-related proteins in our NCBP3 IPs, made us suspect that NCBP3 might generally aid in the formation of 'productive mRNP'. As PHAX and ZC3H18 were previously shown to compete for CBC binding, facilitating snRNA export and exosomal decay, respectively (13), we wondered whether a similar antagonistic relationship would exist between NCBP3 and ZC3H18 for mRNA. We therefore conducted NCBP1-LAP IPs from cell extracts depleted for NCBP3, PHAX or ZC3H18. Depletion of NCBP3 did not notably affect levels of PHAX and ZC3H18 captured by NCBP1-LAP (Figure 6B, compare lanes 8 and 9), which was probably due

to the naturally lower abundance of NCBP3 associating with NCBP1 compared to that of ZC3H18 and PHAX (see Figure 2A, IP condition 12, which was also used in these IPs). However, depletion of ZC3H18 led to a marked increase in levels of captured NCBP3 (Figure 6B, lane 11). This indicates competition between NCBP3 and ZC3H18 for CBC binding, although it is not clear whether such effect is caused by direct competition of the two proteins or whether other unexamined factors are involved. Instead, depletion of PHAX did not change the level of captured NCBP3, and co-depletion of PHAX and ZC3H18 showed a similar effect as depleting ZC3H18 alone (Figure 6B, lanes 10–12). This difference between ZC3H18 and PHAX depletions may be related to the relative cellular amounts of these proteins. The copy number of PHAX in HeLa cells was estimated to be twice that of ZC3H18 (68), potentially leaving

enough PHAX bound to CBC, even upon its partial depletion (as appearing in Figure 6B, lanes 10 and 12), to prevent NCBP3 association. It is also possible that different CBC binding mechanisms are employed by ZC3H18 and PHAX, which result in the different effects of their depletions on NCBP3-NCBP1 interaction. Regardless, these results suggest that NCBP3 competes with ZC3H18 for CBC binding *in vivo*.

## DISCUSSION

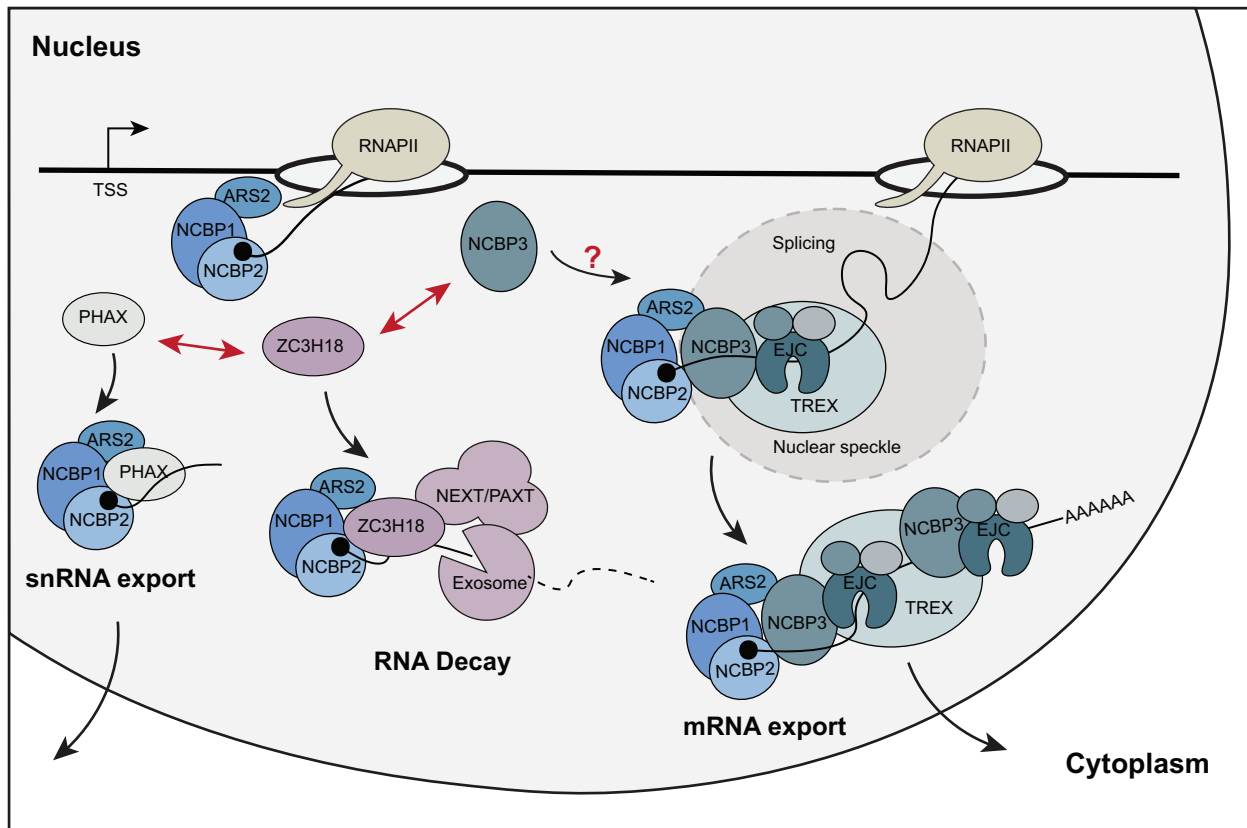
The CBC occupies the nascent RNA cap soon after its emergence from RNAPII, which positions the complex centrally to impact RNA transcription, processing, nuclear transport and decay (5). For two decades, the CBC was known to consist of NCBP1 and NCBP2 (2). However, the previously uncharacterized protein C17orf85 was recently re-named NCBP3 due to its measurable cap-affinity and its suggested ability to form an alternative CBC dimer with NCBP1 (32). Data presented here suggest that the cap-affinity of NCBP3 is unlikely to be relevant for alternative CBC formation. Firstly, NCBP2 depletion prevents, rather than enhances, NCBP3 binding to NCBP1 (Figure 1A). Secondly, NCBP3 does not bind RNA in a cap-proximal manner expected by a CBC component and as displayed by NCBP2 (Figure 4C). Thirdly, NCBP3 associates with both NCBP1 and NCBP2, albeit only weakly (Figure 2A), as reported previously (16,32), making it more likely that NCBP3 interacts with the canonical CBC *in vivo*. Consistently, our interaction profiling of the three NCBPs demonstrated that NCBP3 engages in interaction networks distinct from those of NCBP1 and NCBP2, and instead pointed to the EJC and TREX complexes as its most dominant interactors. NCBP3 localization and functional analyses further strengthened this connection. Based on these data, we therefore propose a model for the involvement of NCBP3 in nuclear RNA metabolism (Figure 7). The CBC-bound RNA is receptive to the binding of different productive and destructive factors, which dynamically exchange during RNP formation to ultimately define transcript fate (31). PHAX and ZC3H18 are established examples of such RNP components, which compete for CBC-binding and ultimately tilt RNA fate towards snRNP/snoRNP assembly or nuclear degradation by the exosome, respectively (13). Based on its abilities to compete with ZC3H18 for NCBP1 binding and to stimulate pA<sup>+</sup> RNA nuclear export (Figure 6), we suggest that NCBP3 is involved in such nuclear RNP definition, favouring EJC-bound transcripts to become export competent.

Our protein-protein interaction profiles provided a direct comparison of the binding partners of the three NCBPs. In contrast to the comparable NCBP1 and NCBP2 interaction profiles, the NCBP3 interaction screening only detected NCBP1 and NCBP2 in low amounts and neither picked up PHAX nor the exosome-related factors ZC3H18, ZCCHC8 and MTR4 (Figure 2A). Instead, NCBP3 abundantly co-purified constituents of the EJC. Curiously though, while most of the EJC and its peripheral components were detected, the EJC core component MLN51 was absent. As an NCBP3-MLN51 interaction was independently established by IP/western blotting analysis (Figure 2B), it is possible

that a technical issue might explain the lack of an MLN51 MS signal, which was also suspiciously absent in previous mRNP IPs (69,70). However, it could also reflect a relatively low abundant MLN51-NCBP3 interaction. On this note, it was suggested that not all of the trimeric EJC (eIF4A3, MAGOH, RBM8A) associates with MLN51, since the protein was substoichiometric to the other EJC core components in both EIF4A3 and MAGOH IPs (71). Moreover, a recent study demonstrated that RNPS1, which was also identified in our NCBP3 interaction screen, and MLN51 associate with the trimeric EJC in a mutually exclusive manner, forming early (mainly nuclear) and late (mainly cytoplasmic) EJCs, respectively (72). This might also help explain why MLN51 is required for NCBP3-EJC interaction *in vitro*, which is not obvious *in vivo*, as the reconstituted environment lacking other proteins, such as RNPS1, that could influence EJC formation. Interactions with MLN51 notwithstanding, the association of the bulk of NCBP3 with EJC was also reflected by their similar RNA binding profiles derived from CLIP datasets; NCBP3 binds close to the splice junctions of multiexonic RNAs, showing two broad peaks at 10–20 nt and 25–35 nt upstream of exon-exon junctions (Figure 4C and Supplemental Figure S4C). This would be consistent with binding around a deposited EJC, which covers around 6–9 nt and a prominent EIF4A3 cross-linking site at ~27 nt upstream of the splice junction (35,42,64,73). Based on this, we speculate that the EJC may direct NCBP3 binding (see below).

Consistent with its physical interaction with EJC proteins, we found NCBP3 co-localizing with nuclear speckles (Figure 4A), which are also enriched for pre-mRNA splicing factors (62). A previous study suggested that pre-mRNA splicing is required for the association of NCBP3 with mRNA *in vitro*, as NCBP3 was absent from intron-depleted mRNPs (70). Further linking NCBP3 to splicing was the establishment of a peripheral association of the protein with the spliceosomal B and C complexes (74). However, notwithstanding these indications, we were not able to establish a role for NCBP3 in alternative splicing, using an EJC-sensitive assay (Figure 5A). We therefore surmise that NCBP3's interaction with the EJC does not impact EJC function. Consistent with this idea, NCBP3 does not require the presence of the EJC for its mRNP association, although this is splicing-dependent (70). Here, we found that NCBP3 requires a fully assembled EJC for *in vitro* interaction (Figure 3B), and that trimeric EJC core integrity is critical for the interaction *in vivo* (Figure 3A). Taking these results together allows us to speculate about the timing of NCBP3 recruitment to the early RNP and its subsequent interaction with the EJC. Trimeric EJC core assembly on RNA occurs after the first step of splicing before exon-exon ligation (75). The EJC-independent RNA binding of NCBP3 (70), would therefore occur between the onset of spliceosome assembly (before EJC recruitment) and until after the first step of splicing (simultaneous with EJC recruitment). Which exact molecular interactions would drive this initial RNA-NCBP3 association remains to be investigated. However, as mentioned above, we suggest that stable EJC formation ultimately provides an anchor point for NCBP3.





**Figure 7.** Model for NCBP3 involvement in RNA fate decisions. Shortly after transcription initiation the CBC (NCBP1 and NCBP2) binds the RNA cap, and forms the CBCA complex with ARS2. Hereafter, the nascent RNA is exposed to different factors, which compete for RNA binding (exemplified by PHAX, ZC3H18 and NCBP3). As shown in the present paper, NCBP3 competes with ZC3H18 for CBC interaction, which is likely tilted in the favour of NCBP3 by its association with EJC and TREX. As such, NCBP3 contributes to licensing pA<sup>+</sup> RNPs for nuclear export. Question mark indicates that the interrelated recruitment mechanisms of NCBP3, EJC and TREX remain to be clarified.

In addition to the EJC, NCBP3 also associates robustly with most known components of the TREX complex, including the complete THO complex (Figure 2A). Previously, NCBP3 was copurified with TREX components ALYREF, DDX39B, THOC2 and CIP29, and it was even considered a putative new TREX subunit (27,58). However, no functional consequence of its association with TREX was reported. Here, we provide evidence for a role of NCBP3 in pA<sup>+</sup> RNA nuclear export by acting redundantly with the TREX component THOC3. Still, we do not consider NCBP3 a central TREX component as its independent depletion does not elicit nuclear accumulation of pA<sup>+</sup> RNA. Instead, NCBP3 may operate tangentially to TREX and this relationship may only be revealed when cells are subjected to the appropriate double depletions of factors. Perhaps the similar phenotype observed upon dual NCBP2 and NCBP3 depletion (32) reflects a related synthetic effect between co-operating RNP factors? In further agreement with an involvement of NCBP3 in the nuclear export of at least some transcripts, we found that NCBP3 positively impacts the expression of a subset of genes. Specifically, upon NCBP3 depletion, mRNAs deriving from multi-exonic transcripts were prone to be downregulated (Figure 5C). We speculate that an abundant number of introns

might increase RNA nuclear residence time and therefore require NCBP3 to fend off nuclear decay activities (as exemplified by its competition with ZC3H18), while providing additional export capacity. As shown previously in both yeast and human cells, efficient nuclear export is required to evade nuclear decay by the RNA exosome (27,43,76). Loss of NCBP3 might therefore de-stabilize these multi-exonic transcripts dually due to their slowed nuclear export and enhanced turnover (Figure 7). NCBP3 might also have more specialized functions. The protein is critical for cells during stress conditions, such as viral infection (32,77); a positive impact of NCBP3 on mRNA biogenesis may therefore also provide support in the expression of stress response genes.

If NCBP3 is not a *bona fide* CBC protein, what is then the functional significance of the m<sup>7</sup>G cap-affinity of the protein? It may bind the cap when CBC is not available at certain undetermined circumstances. Alternatively, and in line with its internal mRNA association, it might potentially contact m<sup>7</sup>G modifications, which are found in mRNAs (78,79), but not in snRNAs and snoRNAs (80). Regardless, and setting the cap affinity of NCBP3 aside, we suggest NCBP3 partakes in dynamic nuclear RNP metabolism, promoting the productive fate of mRNP.

## DATA AVAILABILITY

The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD016038. RNA-seq data have been deposited to the Gene Expression Omnibus (GEO) under the accession number GSE99059.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank the National Center for Dynamic Interactome Research at the Rockefeller University for facility support for the interaction screen, Dr Ina Poser and Prof. Anthony A. Hyman for providing the LAP-tagged cell lines, and Dr Catherine-Laure Tomasetto for providing the MLN51 antibody.

*Author contributions:* Y.D., T.H.J., J.L. and H.L.H. designed the experiments. Y.D., I.B., H.J., C.I., K.R.M., W.M.S. performed the experiments. Y.D., M.S. and K.R.M. performed data analysis. M.S., H.L.H. and S.C. provided critical input. T.H.J. and J.L. supervised the project. Y.D. and T.H.J. wrote the manuscript with input from all co-authors.

## FUNDING

Work in the T.H.J. laboratory was supported by European Research Council Advanced Grant [339953]; Independent Research Fund Denmark and the Lundbeck Foundation; Work in the J.L. laboratory was supported in-part by National Institutes of Health [R01GM126170], and benefitted from the support of the National Center for Dynamic Interactome Research [P41GM109824]; Work in the H.L.H. laboratory was supported by the French Agence Nationale de la Recherche [ANR-13-BSV8-0023, ANR-17-CE12-0021]; Centre National de Recherche Scientifique; Ecole Normale Supérieure and the Institut National de la Santé et de la Recherche Médicale, France. Funding for open access charge: European Research Council [339953].

*Conflict of interest statement.* None declared.

This paper is linked to: [doi.org/10.1093/nar/gkaa743](https://doi.org/10.1093/nar/gkaa743).

## REFERENCES

- Licalatosi,D.D. and Darnell,R.B. (2010) RNA processing and its regulation: global insights into biological networks. *Nat. Rev. Genet.*, **11**, 75–87.
- Izaurrealde,E., Lewis,J., McGuigan,C., Jankowska,M., Darzynkiewicz,E. and Mattaj,I.W. (1994) A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell*, **78**, 657–668.
- Mazza,C., Ohno,M., Segref,A., Mattaj,I.W. and Cusack,S. (2001) Crystal structure of the human nuclear cap binding complex. *Mol. Cell*, **8**, 383–396.
- Calero,G., Wilson,K.F., Ly,T., Rios-Steiner,J.L., Clardy,J.C. and Cerione,R.A. (2002) Structural basis of m7GpppG binding to the nuclear cap-binding protein complex. *Nat. Struct. Biol.*, **9**, 912–917.
- Gonatopoulos-Pournatzis,T. and Cowling,V.H. (2014) Cap-binding complex (CBC). *Biochem. J.*, **457**, 231–242.
- Lenasi,T., Peterlin,B.M. and Barboric,M. (2011) Cap-binding protein complex links pre-mRNA capping to transcription elongation and alternative splicing through positive transcription elongation factor b (P-TEFb). *J. Biol. Chem.*, **286**, 22758–22768.
- Flaherty,S.M., Fortes,P., Izaurrealde,E., Mattaj,I.W. and Gilmartin,G.M. (1997) Participation of the nuclear cap binding complex in pre-mRNA 3' processing. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 11893–11898.
- Narita,T., Yung,T.M.C., Yamamoto,J., Tsuboi,Y., Tanabe,H., Tanaka,K., Yamaguchi,Y. and Handa,H. (2007) NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs. *Mol. Cell*, **26**, 349–365.
- Boulon,S., Verheggen,C., Jady,B.E., Girard,C., Pescia,C., Paul,C., Ospina,J.K., Kiss,T., Matera,A.G., Bordonné,R. et al. (2004) PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. *Mol. Cell*, **16**, 777–787.
- Cheng,H., Dufu,K., Lee,C.-S., Hsu,J.L., Dias,A. and Reed,R. (2006) Human mRNA export machinery recruited to the 5' end of mRNA. *Cell*, **127**, 1389–1400.
- Hosoda,N., Kim,Y.K., Lejeune,F. and Maquat,L.E. (2005) CBP80 promotes interaction of Upf1 with Upf2 during nonsense-mediated mRNA decay in mammalian cells. *Nat. Struct. Mol. Biol.*, **12**, 893–901.
- Andersen,P.R., Domanski,M., Kristiansen,M.S., Storvall,H., Ntini,E., Verheggen,C., Schein,A., Bunkenborg,J., Poser,I., Hallais,M. et al. (2013) The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat. Struct. Mol. Biol.*, **20**, 1367–1376.
- Giacometti,S., Benbahouche,N.E.H., Domanski,M., Robert,M.-C., Meola,N., Lubas,M., Bunkenborg,J., Andersen,J.S., Schulze,W.M., Verheggen,C. et al. (2017) Mutually exclusive CBC-containing complexes contribute to RNA fate. *Cell Reports*, **18**, 2635–2650.
- Müller-McNicoll,M. and Neugebauer,K.M. (2014) Good cap/bad cap: how the cap-binding complex determines RNA fate. *Nat. Struct. Mol. Biol.*, **21**, 9–12.
- Bajczyk,M., Lange,H., Bielewicz,D., Szewc,L., Bhat,S.S., Dolata,J., Kuhn,L., Szweykowska-Kulinska,Z., Gagliardi,D. and Jarmolowski,A. (2020) SERRATE interacts with the nuclear exosome targeting (NEXT) complex to degrade primary miRNA precursors in Arabidopsis. *Nucleic Acids Res.*, **48**, 6839–6854.
- Schulze,W.M., Stein,F., Rettel,M., Nanao,M. and Cusack,S. (2018) Structural analysis of human ARS2 as a platform for co-transcriptional RNA sorting. *Nat. Commun.*, **9**, 1701.
- Hallais,M., Pontvianne,F., Andersen,P.R., Clerici,M., Lener,D., El Houda Benbahouche,N., Gostan,T., Vandermoere,F., Robert,M.-C., Cusack,S. et al. (2013) CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. *Nat. Struct. Mol. Biol.*, **20**, 1358–1366.
- Schulze,W.M. and Cusack,S. (2017) Structural basis for mutually exclusive co-transcriptional nuclear cap-binding complexes with either NELF-E or ARS2. *Nat. Commun.*, **8**, 1302.
- Ohno,M., Segref,A., Bachi,A., Wilm,M. and Mattaj,I.W. (2000) PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell*, **101**, 187–198.
- Pabis,M., Neufeld,N., Steiner,M.C., Bojic,T., Shav-Tal,Y. and Neugebauer,K.M. (2013) The nuclear cap-binding complex interacts with the U4/U6{middle dot}U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. *RNA*, **19**, 1054–1063.
- Le Hir,H., Moore,M.J. and Maquat,L.E. (2000) Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev.*, **14**, 1098–1108.
- Le Hir,H., Saulière,J. and Wang,Z. (2016) The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell Biol.*, **17**, 41–54.
- Heath,C.G., Viphakone,N. and Wilson,S.A. (2016) The role of TREX in gene expression and disease. *Biochem. J.*, **473**, 2911–2935.
- Viphakone,N., Sudbery,I., Griffith,L., Heath,C.G., Sims,D. and Wilson,S.A. (2019) Co-transcriptional loading of RNA export factors shapes the human transcriptome. *Mol. Cell*, **75**, 310–323.
- Gromadzka,A.M., Steckelberg,A.-L., Singh,K.K., Hofmann,K. and Gehring,N.H. (2016) A short conserved motif in ALYREF directs cap- and EJC-dependent assembly of export complexes on spliced mRNAs. *Nucleic Acids Res.*, **44**, 2348–2361.
- Wickramasinghe,V.O. and Laskey,R.A. (2015) Control of mammalian gene expression by selective mRNA export. *Nat. Rev. Mol. Cell Biol.*, **16**, 431–442.

27. Fan, J., Kuai, B., Wu, G., Wu, X., Chi, B., Wang, L., Wang, K., Shi, Z., Zhang, H., Chen, S. *et al.* (2017) Exosome cofactor hMTR4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export. *EMBO J.*, **36**, 2870–2886.
28. Lubas, M., Christensen, M.S., Kristiansen, M.S., Domanski, M., Falkenby, L.G., Lykke-Andersen, S., Andersen, J.S., Dziembowski, A. and Jensen, T.H. (2011) Interaction profiling identifies the human nuclear exosome targeting complex. *Mol. Cell*, **43**, 624–637.
29. Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J.S., Sandelin, A. *et al.* (2016) Identification of a nuclear exosome decay pathway for processed transcripts. *Mol. Cell*, **64**, 520–533.
30. Wolin, S.L. and Maquat, L.E. (2019) Cellular RNA surveillance in health and disease. *Science*, **366**, 822–827.
31. Garland, W. and Jensen, T.H. (2019) Nuclear sorting of RNA. *Wiley Interdiscip Rev RNA*, **21**, e1572.
32. Gebhardt, A., Habjan, M., Benda, C., Meiler, A., Haas, D.A., Hein, M.Y., Mann, A., Mann, M., Habermann, B. and Pichlmair, A. (2015) mRNA export through an additional cap-binding complex consisting of NCBP1 and NCBP3. *Nat. Commun.*, **6**, 8192.
33. Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Lührmann, R. and Stelzl, U. (2012) Dynamic protein-protein interaction wiring of the human spliceosome. *Mol. Cell*, **45**, 567–580.
34. Poser, I., Sarov, M., Hutchins, J.R.A., Hériché, J.-K., Toyoda, Y., Pozniakovskiy, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A.W. *et al.* (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Methods*, **5**, 409–415.
35. Hocq, R., Paternina, J., Alasseur, Q., Genovesio, A. and Le Hir, H. (2018) Monitored eCLIP: high accuracy mapping of RNA-protein interactions. *Nucleic Acids Res.*, **46**, 11553–11565.
36. Hakhverdyan, Z., Domanski, M., Hough, L.E., Oroskar, A.A., Oroskar, A.R., Keegan, S., Dilworth, D.J., Molloy, K.R., Sherman, V., Aitchison, J.D. *et al.* (2015) Rapid, optimized interactomic screening. *Nat. Methods*, **12**, 553–560.
37. Dou, Y., Kalmykova, S., Pashkova, M., Oghbaie, M., Jiang, H., Molloy, K.R., Chait, B.T., Rout, M.P., Fenyö, D., Jensen, T.H. *et al.* (2020) Affinity proteomic dissection of the human nuclear cap-binding complex interactome. *Nucleic Acids Res.*, doi:10.1093/nar/gkaa743.
38. Domanski, M., Molloy, K., Jiang, H., Chait, B., Rout, M., Jensen, T. and LaCava, J. (2012) Improved methodology for the affinity isolation of human protein complexes expressed at near endogenous levels. *BioTechniques*, **1**–6.
39. Cox, J., Hein, M.Y., Lubner, C.A., Paron, I., Nagaraj, N. and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell Proteomics*, **13**, 2513–2526.
40. Smits, A.H., Jansen, P.W.T.C., Poser, I., Hyman, A.A. and Vermeulen, M. (2012) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res.*, **41**, e28.
41. Hulsen, T., de Vlieg, J. and Alkema, W. (2008) BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics*, **9**, 488–486.
42. Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Séraphin, B. and Le Hir, H. (2005) The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat. Struct. Mol. Biol.*, **12**, 861–869.
43. Silla, T., Karadoulama, E., Mąkosa, D., Lubas, M. and Jensen, T.H. (2018) The RNA exosome adaptor ZFC3H1 functionally competes with nuclear export activity to retain target transcripts. *Cell Reports*, **23**, 2199–2210.
44. Iasillo, C., Schmid, M., Yahia, Y., Maqbool, M.A., Descostes, N., Karadoulama, E., Bertrand, E., Andrau, J.-C. and Jensen, T.H. (2017) ARS2 is a general suppressor of pervasive transcription. *Nucleic Acids Res.*, **45**, 10229–10241.
45. Winczura, K., Schmid, M., Iasillo, C., Molloy, K.R., Harder, L.M., Andersen, J.S., LaCava, J. and Jensen, T.H. (2018) Characterizing ZC3H18, a multi-domain protein at the interface of RNA production and destruction decisions. *Cell Reports*, **22**, 44–58.
46. Silla, T., Schmid, M., Dou, Y., Garland, W., Milek, M., Imami, K., Johnsen, D., Polak, P., Andersen, J.S., Selbach, M. *et al.* (2020) The human ZC3H3 and RBM26/27 proteins are critical for PAXT-mediated nuclear RNA decay. *Nucleic Acids Res.*, **48**, 2518–2530.
47. Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.
48. Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
49. Zhu, Y., Xu, G., Yang, Y.T., Xu, Z., Chen, X., Shi, B., Xie, D., Lu, Z.J. and Wang, P. (2019) POSTAR2: deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res.*, **47**, D203–D211.
50. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F. and Manke, T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.*, **44**, W160–W165.
51. Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M. *et al.* (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell*, **46**, 674–690.
52. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, U., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
53. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
54. Hubner, N.C., Bird, A.W., Cox, J., Spletstoeser, B., Bandilla, P., Poser, I., Hyman, A. and Mann, M. (2010) Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J. Cell Biol.*, **189**, 739–754.
55. Degot, S., Le Hir, H., Alpy, F., Kedinger, V., Stoll, I., Wendling, C., Séraphin, B., Rio, M.-C. and Tomasetto, C. (2004) Association of the breast cancer protein MLN51 with the exon junction complex via its speckle localizer and RNA binding module. *J. Biol. Chem.*, **279**, 33702–33715.
56. Dagueuet, E., Baguet, A., Degot, S., Schmidt, U., Alpy, F., Wendling, C., Spiegelhalter, C., Kessler, P., Rio, M.-C., Le Hir, H. *et al.* (2012) Perispeckles are major assembly sites for the exon junction core complex. *Mol. Biol. Cell*, **23**, 1765–1782.
57. Wang, Z., Murigneux, V. and Le Hir, H. (2014) Transcriptome-wide modulation of splicing by the Exon Junction Complex. *Genome Biol.*, **15**, 551.
58. Dufu, K., Livingstone, M.J., Seebacher, J., Gygi, S.P., Wilson, S.A. and Reed, R. (2010) ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. *Genes Dev.*, **24**, 2043–2053.
59. Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N. and Reed, R. (2005) Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev.*, **19**, 1512–1517.
60. Kota, K.P., Wagner, S.R., Huerta, E., Underwood, J.M. and Nickerson, J.A. (2008) Binding of ATP to UAP56 is necessary for mRNA export. *J. Cell. Sci.*, **121**, 1526–1537.
61. Dias, A.P., Dufu, K., Lei, H. and Reed, R. (2010) A role for TREX components in the release of spliced mRNA from nuclear speckle domains. *Nat. Commun.*, **1**, 97.
62. Spector, D.L. and Lamond, A.I. (2011) Nuclear speckles. *Cold Spring Harb. Perspect. Biol.*, **3**, a000646.
63. Saulière, J., Murigneux, V., Wang, Z., Marquet, E., Barbosa, I., Le Tonquèze, O., Audic, Y., Paillard, L., Crollius, H.R. and Le Hir, H. (2012) CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat. Struct. Mol. Biol.*, **19**, 1124–1131.
64. Le Hir, H., Izaurralde, E., Maquat, L.E. and Moore, M.J. (2000) The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.*, **19**, 6860–6869.
65. Boehm, V., Britto-Borges, T., Steckelberg, A.-L., Singh, K.K., Gerbracht, J.V., Gueney, E., Blazquez, L., Altmüller, J., Dieterich, C. and Gehring, N.H. (2018) Exon Junction Complexes Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity. *Mol. Cell*, **72**, 482–495.
66. Chi, B., Wang, Q., Wu, G., Tan, M., Wang, L., Shi, M., Chang, X. and Cheng, H. (2013) Aly and THO are required for assembly of the



- human TREX complex and association of TREX components with the spliced mRNA. *Nucleic Acids Res.*, **41**, 1294–1306.
67. Hautbergue, G.M., Hung, M.-L., Walsh, M.J., Snijders, A.P.L., Chang, C.-T., Jones, R., Ponting, C.P., Dickman, M.J. and Wilson, S.A. (2009) UIF, a New mRNA export adaptor that works together with REF/ALY, requires FACT for recruitment to mRNA. *Curr. Biol.*, **19**, 1918–1924.
68. Bekker-Jensen, D.B., Kelstrup, C.D., Batth, T.S., Larsen, S.C., Haldrup, C., Bramsen, J.B., Sørensen, K.D., Høyer, S., Ørntoft, T.F., Andersen, C.L. *et al.* (2017) An Optimized shotgun strategy for the rapid generation of comprehensive human proteomes. *Cell Syst.*, **4**, 587–599.
69. Tange, T.Ø., Shibuya, T., Jurica, M.S. and Moore, M.J. (2005) Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA*, **11**, 1869–1883.
70. Merz, C., Urlaub, H., Will, C.L. and Lührmann, R. (2007) Protein composition of human mRNPs spliced in vitro and differential requirements for mRNP protein recruitment. *RNA*, **13**, 116–128.
71. Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z. and Moore, M.J. (2012) The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell*, **151**, 750–764.
72. Mabin, J.W., Woodward, L.A., Patton, R.D., Yi, Z., Jia, M., Wysocki, V.H., Bundschuh, R. and Singh, G. (2018) The exon junction complex undergoes a compositional switch that alters mRNP structure and nonsense-mediated mRNA decay activity. *Cell Rep.*, **25**, 2431–2446.
73. Andersen, C.B.F., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L.P., Pedersen, J.S., Séraphin, B., Le Hir, H. and Andersen, G.R. (2006) Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science*, **313**, 1968–1972.
74. Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H. and Lührmann, R. (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature*, **452**, 846–850.
75. Gehring, N.H., Lamprinak, S., Hentze, M.W. and Kulozik, A.E. (2009) The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. *PLoS Biol.*, **7**, e1000120.
76. Tudek, A., Schmid, M., Makaras, M., Barrass, J.D., Beggs, J.D. and Jensen, T.H. (2018) A nuclear export block triggers the decay of newly synthesized polyadenylated RNA. *Cell Rep.*, **24**, 2457–2467.
77. Gebhardt, A., Bergant, V., Schnepf, D., Moser, M., Meiler, A., Togbe, D., Mackowiak, C., Reinert, L.S., Paludan, S.R., Ryffel, B. *et al.* (2019) The alternative cap-binding complex is required for antiviral defense in vivo. *PLoS Pathog.*, **15**, e1008155.
78. Chu, J.-M., Ye, T.-T., Ma, C.-J., Lan, M.-D., Liu, T., Yuan, B.-F. and Feng, Y.-Q. (2018) Existence of internal N7-methylguanosine modification in mRNA determined by differential enzyme treatment coupled with mass spectrometry analysis. *ACS Chem. Biol.*, **13**, 3243–3250.
79. Zhang, L.-S., Liu, C., Ma, H., Dai, Q., Sun, H.-L., Luo, G., Zhang, Z., Zhang, L., Hu, L., Dong, X. *et al.* (2019) Transcriptome-wide mapping of internal N7-methylguanosine methylome in mammalian mRNA. *Mol. Cell*, **74**, 1304–1316.
80. Enroth, C., Poulsen, L.D., Iversen, S., Kirpekar, F., Albrechtsen, A. and Vinther, J. (2019) Detection of internal N7-methylguanosine (m7G) RNA modifications by mutational profiling sequencing. *Nucleic Acids Res.*, **47**, e126.