Mutually Exclusive CBC-Containing Complexes Contribute to RNA Fate

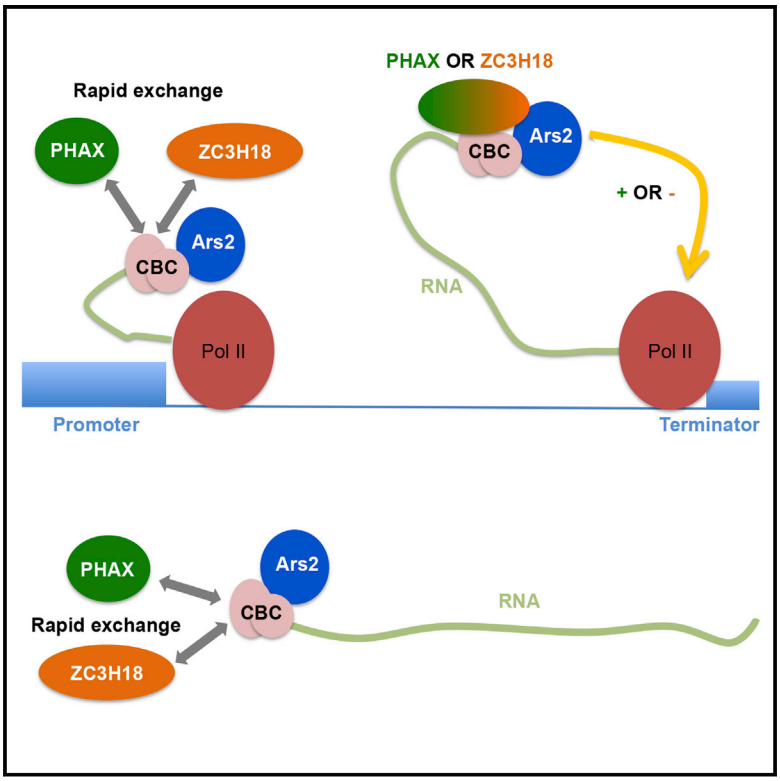
Simone Giacometti, Nour El Houda Benbahouche, Michal Domanski, Marie-Cécile Robert, Nicola Meola, Michal Lubas, Jakob Bukenborg, Jens B. Andersen, Wiebke Schulze, Céline Verheggen, et al.

Article

Cell Reports

Mutually Exclusive CBC-Containing Complexes Contribute to RNA Fate

Graphical Abstract



Highlights

- PHAX and ZC3H18 compete for binding to the nuclear CBC
- PHAX and ZC3H18 have opposite effects on the fate of snRNA precursors and other RNAs
- PHAX, ARS2, and ZC3H18 bind capped RNAs without strong preference for given transcripts
- CBC-containing complexes are short lived in vivo, with a lifetime of a few seconds

Authors

Simone Giacometti, Nour El Houda Benbahouche, Michal Domanski, ..., Grzegorz Kudla, Torben Heick Jensen, Edouard Bertrand

Correspondence

gkudla@gmail.com (G.K.), thj@mbg.au.dk (T.H.J.), edouard.bertrand@igmm.cnrs.fr (E.B.)

In Brief

The nuclear CBC plays diverse roles in RNA biogenesis and it is not clear how selective effects are achieved for individual RNA families. Giacometti et al. suggest that RNA fate involves the formation of short-lived, mutually exclusive CBC complexes, which may only be consequential at particular checkpoints during RNA biogenesis.

Accession Numbers

GSE94427

Giacometti et al., 2017, Cell Reports *18*, 2635–2650
March 14, 2017
http://dx.doi.org/10.1016/j.celrep.2017.02.046

Cell Reports
Article

Mutually Exclusive CBC-Containing Complexes Contribute to RNA Fate

Simone Giacometti,[1,2,3,6,7] Nour El Houda Benbahouche,[2,6] Michal Domanski,[1,6,8] Marie-Cécile Robert,[2] Nicola Meola,[1] Michal Lubas,[1,9] Jakob Bukenborg,[4] Jens S. Andersen,[4] Wiebke M. Schulze,[5] Celine Verheggen,[2] Grzegorz Kudla,[3,*] Torben Heick Jensen,[1,*] and Edouard Bertrand[2,10,*]

[1]Centre for mRNP Biogenesis and Metabolism, Department of Molecular Biology and Genetics, Aarhus University, C. F. Møllers Allé 3, Bldg. 1130, 8000 Aarhus C, Denmark
[2]Unité Mixte de Recherche 5535, Institut de Génétique Moléculaire de Montpellier, CNRS and Montpellier University, 34293 Montpellier, France
[3]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4, UK
[4]Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark
[5]Grenoble Outstation, European Molecular Biology Laboratory, 71 Avenue des Martyrs, CS 90181, 38042 Grenoble, France
[6]Co-first author
[7]Present address: Diabetes Center, UCSF School of Medicine, 513 Parnassus Avenue, San Francisco, CA 94143, USA
[8]Present address: Department of Chemistry and Biochemistry, Freiestrasse 3, University of Bern, 3012 Bern, Switzerland
[9]Present address: Biotech Research and Innovation Centre, University of Copenhagen, 2200 Copenhagen N, Denmark
[10]Lead Contact
*Correspondence: gkudla@gmail.com (G.K.), thj@mbg.au.dk (T.H.J.), edouard.bertrand@igmm.cnrs.fr (E.B.)
http://dx.doi.org/10.1016/j.celrep.2017.02.046

SUMMARY

The nuclear cap-binding complex (CBC) stimulates processing reactions of capped RNAs, including their splicing, 3′-end formation, degradation, and transport. CBC effects are particular for individual RNA families, but how such selectivity is achieved remains elusive. Here, we analyze three main CBC partners known to impact different RNA species. ARS2 stimulates 3′-end formation/transcription termination of several transcript types, ZC3H18 stimulates degradation of a diverse set of RNAs, and PHAX functions in pre-small nuclear RNA/small nucleolar RNA (pre-snRNA/snoRNA) transport. Surprisingly, these proteins all bind capped RNAs without strong preferences for given transcripts, and their steady-state binding correlates poorly with their function. Despite this, PHAX and ZC3H18 compete for CBC binding and we demonstrate that this competitive binding is functionally relevant. We further show that CBC-containing complexes are short lived in vivo, and we therefore suggest that RNA fate involves the transient formation of mutually exclusive CBC complexes, which may only be consequential at particular checkpoints during RNA biogenesis.

INTRODUCTION

All RNA polymerase II (RNAPII) transcripts undergo processing events that are essential for their function. Early during RNA synthesis, an m7-G cap is added to the nascent 5′ end by an enzymatic complex that binds the serine 5 phosphorylated form of the C-terminal domain (CTD) of RNAPII (Bentley, 2014). By protecting the nascent RNA from 5′ to 3′ degradation, the cap thus represents the hallmark of a successfully initiated RNAPII transcript. Importantly, the cap also serves a key role in many aspects of nuclear RNA biology (Gonatopoulos-Pournatzis and Cowling, 2014). Nuclear cap functions are mediated by the CBP80 and CBP20 proteins (also named NCBP1 and NCBP2), composing the nuclear cap-binding complex (CBC) that associates co-transcriptionally with the nascent RNA (Glover-Cutter et al., 2008; Görnemann et al., 2005; Narita et al., 2007). CBP20 interacts directly with the m7-G cap through its classical RNA recognition motif (RRM), while CBP80 ensures high-affinity binding of the full CBC and provides a platform for interactions with other factors (Izaurralde et al., 1994; Calero et al., 2002; Mazza et al., 2002).

The CBC is highly specific for guanosine caps modified at position N7 (m7-G cap). Cap-adjacent nucleotides may also carry modifications, but it is believed that these nucleotides increase CBC affinity in a rather non-sequence-specific manner (Worch et al., 2005). In the following, we therefore refer to "capped RNA" as transcripts carrying an m7-G cap, regardless of the identity or modification of the adjacent nucleotides. The CBC is believed to bind all classes of m7-G-capped RNAs, including precursors and mature forms of mRNAs, stable long non-coding RNAs (lncRNAs), non-adenylated histone RNAs, and precursors of spliceosomal small nuclear RNAs (snRNAs). It also associates with m7-G capped forms of small nucleolar RNAs (snoRNAs) and labile lncRNAs, such as promoter upstream transcripts (PROMPTs; Preker et al. 2008). Through its cap association, the CBC affects nuclear RNA metabolism in ways that appear specific for different RNA families. In the case of conventional mRNAs, the CBC stimulates the splicing of cap-proximal introns, the processing of RNA 3′ ends, and the formation of

export-competent ribonucleoproteins (RNPs) (Cheng et al., 2006; Flaherty et al., 1997; Izaurralde et al., 1994). Stimulation of RNA splicing and export has been proposed to involve interactions of the CBC with the U4/U6.U5 tri-small nuclear RNP (snRNP) and ALYREF, respectively (Cheng et al., 2006; Pabis et al., 2013). In the case of non-adenylated histone mRNAs, the CBC promotes their 3′ end formation in a process involving interactions with the ARS2, NELF-E, and SLBP proteins (Gruber et al., 2012; Hallais et al., 2013; Narita et al., 2007). In the case of PROMPTs and other short-lived transcripts, such as products of readthrough transcription, the CBC recruits ARS2, ZC3H18, and the nuclear exosome targeting (NEXT) complex, composed of RBM7, ZCCHC8, and hMTR4 (Lubas et al., 2011). This leads to the formation of the CBC-NEXT (CBCN) complex (Figure 1A), which promotes RNA degradation via the nuclear RNA exosome (Andersen et al., 2013; Lubas et al., 2015). Finally, in the case of snRNAs, the CBC promotes transcription termination, aided by ARS2, and nuclear export of the resulting precursors (Andersen et al., 2013; Hallais et al., 2013; Ohno et al., 2000). The latter activity involves the so-called CBC-ARS2-PHAX (CBCAP) complex (Hallais et al., 2013; Figure 1A), where PHAX acts as an adaptor between the CBC/RNP complex and the nuclear export receptor CRM1 (Ohno et al., 2000). PHAX and the CBC are also involved in the biogenesis of capped snoRNAs, directing the intranuclear transport of nascent snoRNAs to Cajal bodies (Boulon et al., 2004).

Such a broad collection of CBC functions raises the question of how specificity is achieved; that is, how are different RNA families identified and brought to their proper processing machineries? This question is particularly relevant, given the dual RNA-productive and RNA-degradative effects imposed by the CBC on nuclear RNA (Andersen et al., 2013; Hallais et al., 2013). At least part of the answer lies in the different protein partners of the CBC complex (Müller-McNicoll and Neugebauer, 2014). As mentioned above, distinct CBC effectors drive different processing reactions, and their recognition of particular RNA families, or even individual transcripts, could potentially provide specificity. This concept is supported by studies of snRNAs and mRNAs in *Xenopus* oocytes, which indicate that the protein composition of the corresponding capped RNPs is determined by the RNA length and intronic content (Masuyama et al., 2004; Ohno et al., 2002). On the one hand, introns lead to the deposition of the exon junction complex (EJC) onto spliced RNAs (Ideue et al., 2007; Le Hir et al., 2000a), and the EJC communicates with the CBC to recruit the mRNA export adaptor ALY/REF (Cheng et al., 2006). On the other hand, RNA length appears to determine whether PHAX efficiently associates with CBC-bound RNAs or not (Masuyama et al., 2004; Ohno et al., 2002). Indeed, PHAX was suggested to specifically associate with short RNAs due to its active exclusion by hnRNPC tetramers, which bind selectively to RNAs longer than 200 nt (McCloskey et al., 2012). Whether this mechanism applies to all nuclear RNAs is currently unknown. How other CBC effectors discriminate their transcript targets and how effector-target recognition translates into biological activity are also unanswered questions.

In this study, we employ transcriptome-wide in vivo RNA cross-linking methodology, protein-protein interaction assays, factor depletions followed by substrate analysis, and fluorescence microscopy to functionally characterize three key CBC partners: ARS2, PHAX, and ZC3H18. Surprisingly, we find that the target specificities of these factors at steady state are rather broad and therefore unable to explain the RNA family-specific activities of the CBC. In contrast, our data suggest a model where short-lived, mutually exclusive CBC-containing complexes determine RNA fate by reacting to molecular cues imposed at specific time points during RNA biogenesis.

RESULTS

ARS2, PHAX, and ZC3H18 Bind mRNA/pre-mRNA in a Cap-Proximal Fashion

To characterize how CBC-interacting factors with different biological activities might achieve RNA family-specific effects, we first performed individual-nucleotide resolution UV cross-linking and immunoprecipitation (iCLIP) with ARS2, PHAX, and ZC3H18. These proteins all bind RNA and associate with the CBC, but with distinct outcomes, providing good models to test whether substrate selectivity is accomplished by the specific recognition of RNA by CBC partners. As comparisons, we conducted iCLIP with CBP20, providing a useful baseline on which to compare CBC partners, and included our previous iCLIP analysis of the NEXT component RBM7 (Lubas et al., 2015).

For all proteins except ZC3H18, iCLIP was performed using HeLa Kyoto cell lines expressing, under the control of the respective endogenous gene promoters, localization and affinity purification (LAP)-tagged proteins with an N- or C-terminal GFP moiety (Andersen et al., 2013; Figure S1A). Since a tagged ZC3H18 HeLa Kyoto cell line could not be obtained, we instead employed a C-terminally 3xFLAG-tagged ZC3H18 cDNA, which was introduced in a single copy into HEK293 Flp-In T-REx cells (Andersen et al., 2013). All interrogated factors could be efficiently cross-linked to RNA in a UV-dependent manner and extensive RNase I treatment of immunoprecipitated (IPed) material confirmed that the majority of RNA was attached to the relevant proteins (Figure S1B). The "no-tag" control cell lines yielded no detectable PCR products (Figure S1C), implying a low experimental background. Each immunoprecipitation (IP) iCLIP library was produced in duplicate (Table S1) and the distribution of total mapped reads was calculated (Table S2). The replicates were generally similar to each other and different from both cytoplasmic poly(A)$^+$ RNAs and rRNA-depleted total RNAs, revealing both reproducibility and specificity (Figures 1B and S1D; Table S2).

As expected from their CBC connections (Andersen et al., 2013; Hallais et al., 2013), ARS2, PHAX, ZC3H18, and RBM7 mainly bound to capped RNAs (Figure 1B). CBP20 was highly enriched on "mRNA first exons" (Table S2), in line with its direct binding to the cap. ARS2 and PHAX were both enriched on snRNAs and capped snoRNAs, consistent with their functions in snRNA biogenesis. However, all interrogated factors bound mRNA as their primary transcript biotype. For PHAX, this was somewhat unexpected, given its reported absence from long capped transcripts in *Xenopus* oocytes (Masuyama et al., 2004; Ohno et al., 2002). Selected iCLIP substrates were, however, validated by regular IPs followed by RNase protection or qRT-PCR analyses (Figures S2A–S2C), as well as by manual cross-linking and immunoprecipitation (CLIP) experiments (Figure S2D).

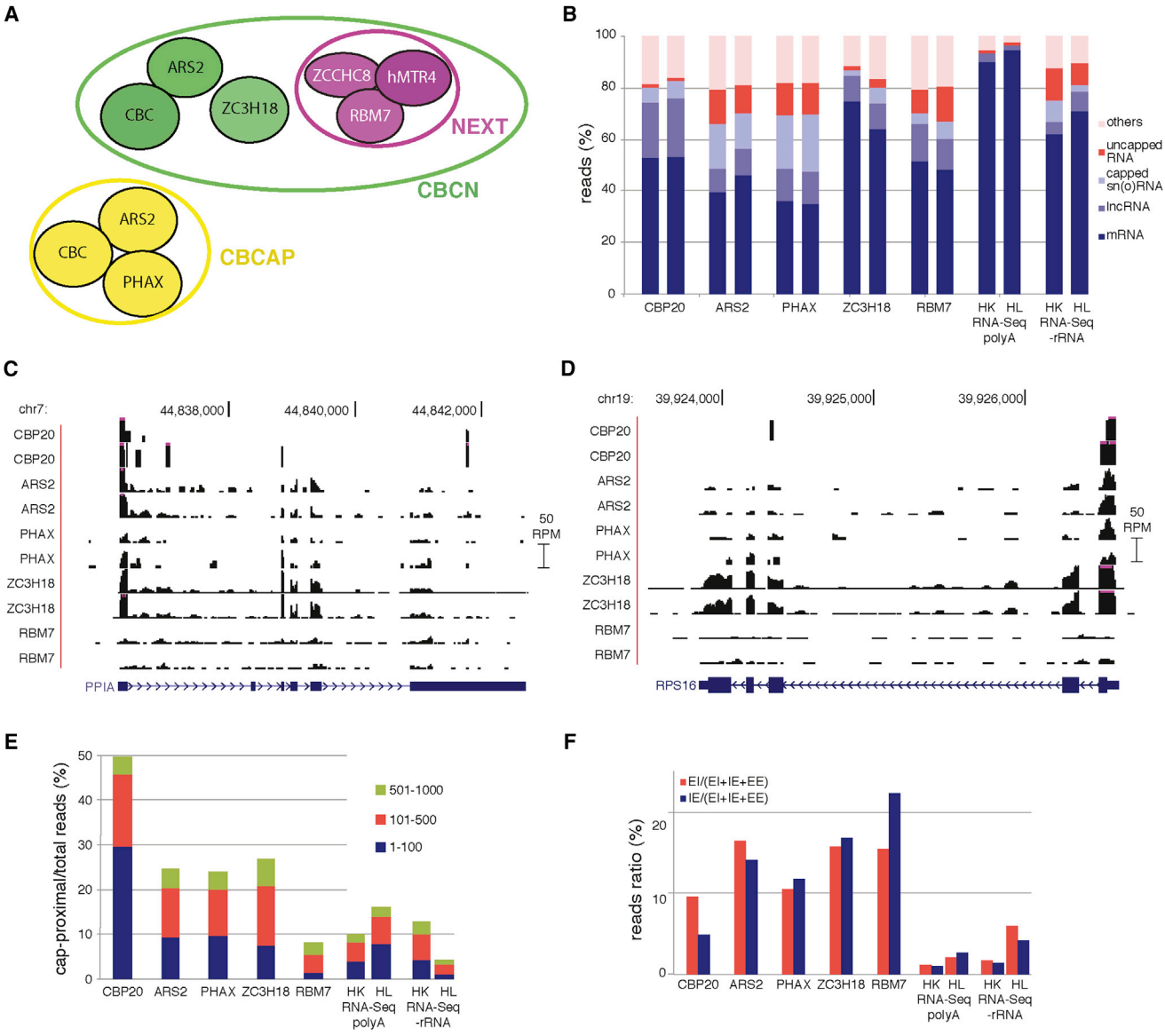


Figure 1. Cap-Proximal mRNA Binding by ARS2, PHAX, and ZC3H18

(A) Schematic overview of the different protein complexes relevant for this study. CBCAP is shown in yellow, NEXT is in purple, and CBCN is circled in green. See the text for details.

(B) Fractions of iCLIP reads, from replicate libraries, mapping to the indicated classes of capped or uncapped RNA expressed as proportions of total library reads. Reads marked as "others" could not be unambiguously assigned to any of the above categories. For comparison, we show cytoplasmic poly(A)$^+$-selected and rRNA-depleted RNA-seq data from HEK293 (HK) and HeLa (HL) cells.

(C and D) Genome browser views of representative protein-coding genes *PPIA* (C) and *RPS16* (D), showing iCLIP reads from replicate CBP20, ARS2, PHAX, ZC3H18, and RBM7 samples. Reads mapping to the *PPIA* and *RPS16* RNAs are shown as mapped reads per million (RPM) library reads (see scale bar to the right of the image). Purple color implies that displayed reads exceed the scale used.

(E) Fractions of iCLIP or RNA-seq reads mapping within cap-proximal regions of 100, 500, or 1,000 nt of 5,769 well-annotated pre-mRNA genes. The iCLIP results represent averages of replicate experiments.

(F) Fractions of exon-intron (EI) and intron-exon (IE) junction reads, averaged between replicate experiments, mapping over RefSeq pre-mRNAs. Fractions were calculated as EI/(EI + IE + EE) and IE/(EI + IE + EE), as indicated. Note that EI fractions are higher than IE fractions for CBP20 libraries in agreement with the cap-binding nature of this protein. Conversely, IE fractions are higher than EI fractions for RBM7 libraries, as previously reported (Lubas et al., 2015). EE, exon-exon junction reads.

Visual examination of representative examples of canonical pre-mRNAs demonstrated that CBP20, ARS2, PHAX, and ZC3H18 exhibited a cap-proximal cross-linking preference (Figures 1C and 1D). Although such tendency was also reported for RBM7 (Lubas et al., 2015), this protein associated relatively more with the bodies of the examined transcripts. To more generally assess

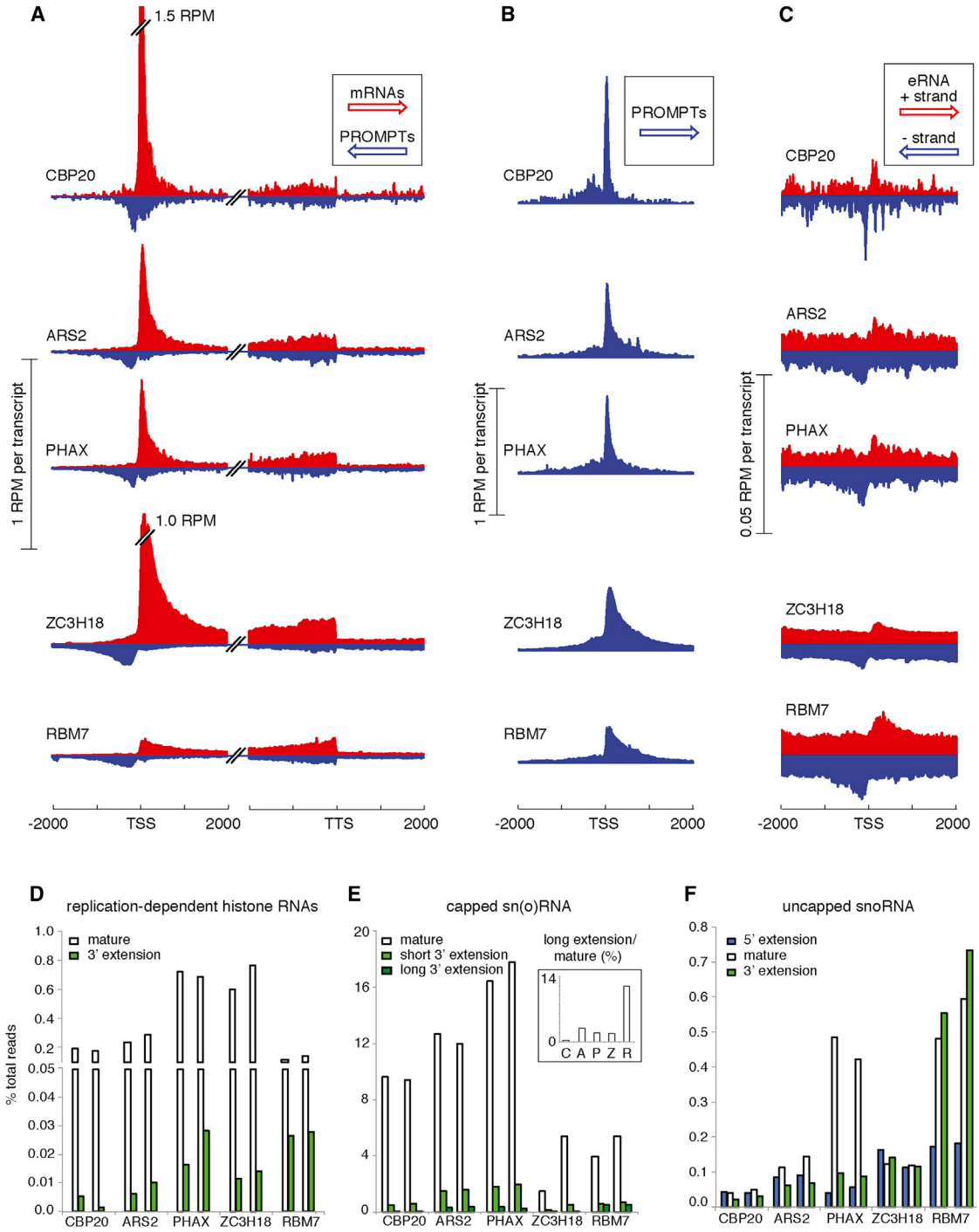


Figure 2. ARS2, PHAX, and ZC3H18 Are Targeted to Common RNA Families

(A) Density profiles of reads from the indicated iCLIP libraries displayed as reads per million (RPM) library reads, around ±2-kb regions of transcription start sites (TSSs; left part) and transcript termination sites (TTSs; right part) of the protein-coding genes from Figure 1E. Transcription directions are indicated by arrows as forward (mRNA direction) and reverse (PROMPT direction). Red and blue reads map to forward and reverse strands, respectively. Signal corresponding to 1 RPM is indicated. Note that CBP20 and ZC3H18 mRNA profiles were disrupted to ease visual inspection.

(B) Density profiles as in (A) but only showing reverse read densities in ±2-kb regions anchored around PROMPT TSSs as defined by CAGE summits (Chen et al., 2016). Signal corresponding to 1 RPM is indicated.

(C) Density profiles as in (A) but showing forward and reverse read densities in ±2-kb regions anchored around eRNA TSSs as defined by CAGE summits (Chen et al., 2016). Signal corresponding to 0.05 RPM is indicated.

(legend continued on next page)

factor binding, we employed a set of 5,769 well-annotated pre-mRNAs, containing no other annotated transcription start sites (TSSs) or transcript termination sites (TTSs) in the interrogated regions, and we calculated the fraction of iCLIP reads falling within the first 100, 500, or 1,000 cap-proximal nucleotides. As expected, the CBP20 CLIP signal was highly enriched at cap-proximal positions (Figure 1E) and consistent with the individually examined pre-mRNAs, ARS2, PHAX, and ZC3H18 displayed more frequent cap-proximal reads than RBM7 or than that observed by the distribution of RNA sequencing (RNA-seq) reads, using either cytoplasmic poly(A)$^+$ RNAs or rRNA-depleted total RNAs.

To examine the maturation status of mRNAs bound by CBP20, ARS2, PHAX, and ZC3H18, we next calculated the fraction of exon-intron (EI) or intron-exon (IE) junction reads in the respective libraries. Whereas RNA-seq datasets contained mostly spliced reads, iCLIP with CBC and its binding partners recovered many unspliced transcripts, consistent with the nuclear localization of the proteins (Figure 1F). CBP20 was most strongly enriched on spliced species, closely followed by PHAX, ARS2, and ZC3H18 (Figure 1F). As expected, RBM7 exhibited a relatively stronger binding to IE junctions, consistent with its accumulation in the 3′ ends of introns (Lubas et al., 2015).

Taking these analyses together, we conclude that CBP20, ARS2, PHAX, and ZC3H18 associate with both immature and mature mRNAs with a common preference for cap-proximal binding, consistent with previous biochemical experiments (Andersen et al., 2013; Hallais et al., 2013; Izaurralde et al., 1992; Ohno et al., 1987). RBM7, on the other hand, associates with RNA in a less cap-proximal fashion. Hence, besides the surprising interaction of PHAX with pre-mRNA/mRNA, we note that the distinct ZC3H18 and RBM7 binding profiles suggest that a stable CBCN complex does not readily form within nuclear pre-mRNP/mRNP.

Targeting of ARS2, PHAX, and ZC3H18 to Different Classes of RNAPII-Derived Transcripts

To further characterize transcript association of the investigated factors, we first generated metagene profiles of read densities from individual libraries by anchoring sequence tags to pre-mRNA TSSs or TTSs. As expected from our previous analyses, this revealed sharp cap-proximal peaks of CBP20, ARS2, PHAX, and ZC3H18, as well as a more moderate enrichment of RBM7 (Figure 2A, red coloring). No major differences were observed for these proteins near the RNA 3′ ends. Cap-proximal binding profiles for CBP20, ARS2, PHAX, and ZC3H18 were also apparent for reverse-transcribed PROMPTs (Figure 2A, blue coloring), which became clearer when CLIP signals were anchored to PROMPT 5′ ends (Figure 2B) as defined by cap analysis of gene expression (CAGE) data (Ntini et al., 2013). As for pre-mRNAs, RBM7 bound PROMPTs with a more moderate cap-proximal tendency. Interrogated proteins also accumulated close to the cap of long intergenic non-coding RNAs (lincRNAs; Figure S3) and enhancer RNAs (eRNAs; Figure 2C), although the low-abundant nature of the latter in the utilized exosome-proficient cells only allowed a moderate spatial signal resolution.

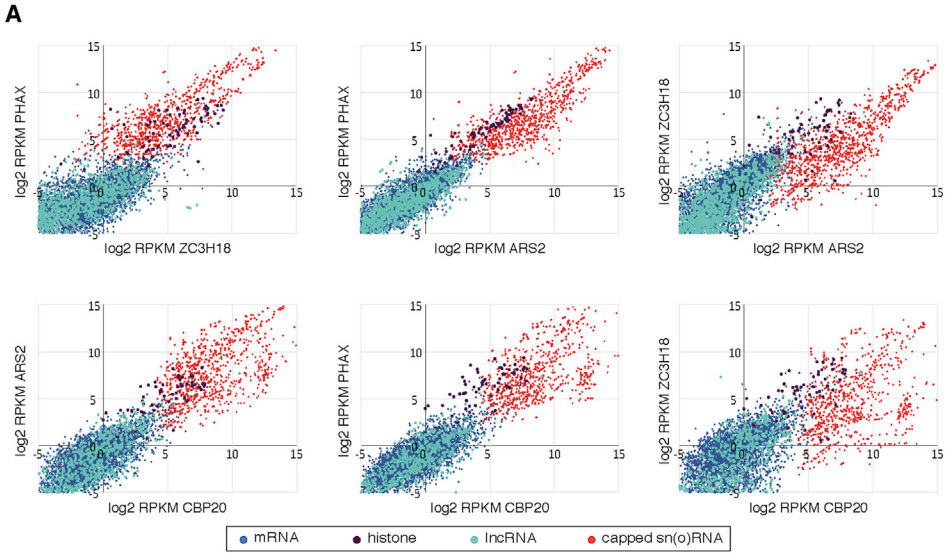
We next examined binding of factors to replication-dependent histone (RDH) RNAs, which are 3′ end processed by U7 snRNA and therefore not polyadenylated. All of the investigated proteins bound to histone mRNAs, with PHAX and ZC3H18 showing the highest fractions of CLIP reads (Figure 2D). RDH genes also generate 3′-extended transcripts that may terminate at cryptic downstream polyadenylation (pA) sites (Gruber et al., 2012). Estimating iCLIP reads mapping to such 3′ extensions relative to mature RDH transcript revealed elevated RBM7 binding compared to the other factors (Figure 2D). A similar tendency was also observed when interrogating independently transcribed sn(o)RNAs (Figure 2E, inset). Primary snRNA transcripts are cleaved by the Integrator complex to generate pre-snRNAs carrying extensions of less than 20 nt ("short 3′ extensions"), which are exported to the cytoplasm by CBC and PHAX to be processed into mature trimethyl guanosine (TMG) capped snRNAs (Ohno et al., 2000). snRNA genes also produce transcripts carrying 3′ extensions of a few hundred nucleotides ("long 3′ extensions") and whose degradation relies on ZC3H18 and NEXT (Andersen et al., 2013). Consistently, RBM7 binding was again elevated on long 3′ extensions relative to mature RNAs (Figure 2E, inset), but somewhat surprisingly this was not the case for ZC3H18 (see below). Finally, binding of factors to snoRNAs deriving from splicing of their host introns was analyzed and revealed robust RBM7 binding to mature snoRNAs and their 3′ extensions (Figure 2F), consistent with NEXT-mediated decay from intronic 3′ ends (Lubas et al., 2015). Interestingly, PHAX bound strongly to mature uncapped snoRNAs, whereas CBP20 and ARS2 did not, suggesting that PHAX may be recruited to these RNAs independently of CBC/ARS2.

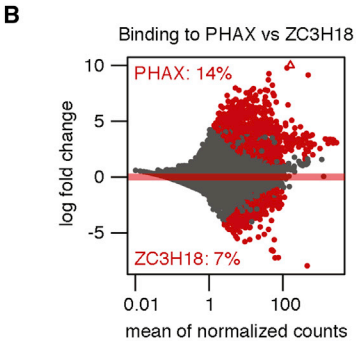
Taking the data together, we conclude that the CBC and its partners generally bind the same families of coding and non-coding capped RNAs. However, some differences can be observed. First, RBM7 contacts unprocessed, long 3′ extended snRNA and RDH transcripts, which most likely mirrors the NEXT-mediated activity of the RNA exosome on these species. Second, ARS2 and PHAX display a moderate enrichment on snRNAs as compared to CBP20, for example, which is consistent with their role in snRNA export. This is, however, contrasted by their quantitatively robust binding to mRNA (Table S2). Such limited specificity of ARS2 and PHAX for snRNAs appears insufficient to faithfully identify these RNAs within the nucleus.

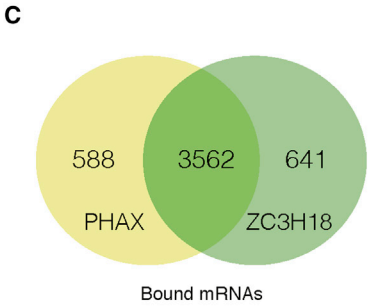
ARS2, PHAX, and ZC3H18 Display Limited Specificity within Separate RNA Families

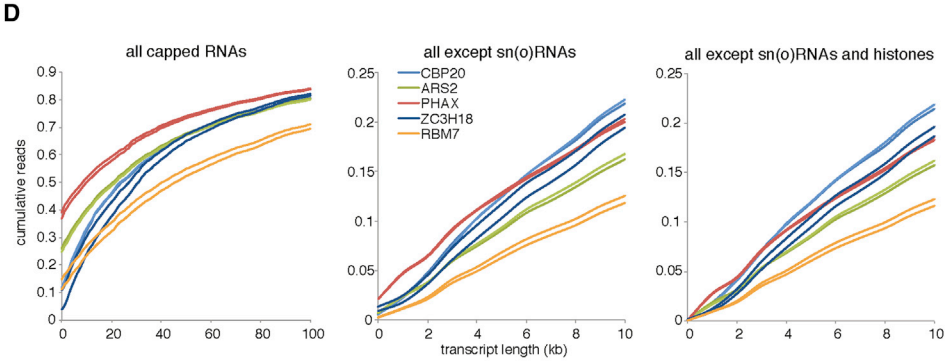
Although ARS2, PHAX, and ZC3H18 bind families of capped RNA without strong selectivity, they might still bind different

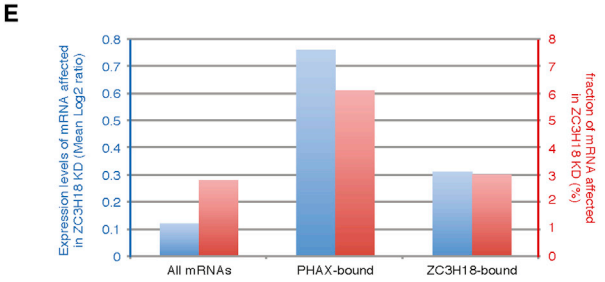
(D) Proportion of reads from the indicated replicate libraries mapping to mature (white columns) and 3′ extended regions (light green columns) of RDH RNAs. "3′ extensions" denote 1–500 nt downstream of the annotated mature RDH RNA 3′ end. Note disruption of the y axis to ease visual inspection of all data.
(E) Proportion of reads mapping to mature (white columns), short (light green), and long (dark green) 3′ extended regions of independently transcribed sn(o)RNAs. "Short 3′ extensions" and "long 3′ extensions" denote 1–20 nt and 50–500 nt, respectively, downstream of the annotated mature sn(o)RNA 3′ ends. The inset shows the ratio of reads mapping to long 3′ extensions relative to mature RNA.
(F) Proportion of reads mapping to 5′extended- (blue columns), mature- (white columns) and 3′extension- (light green columns) regions of uncapped snoRNAs located in introns. 5′- and 3′-extension denote regions from the mature snoRNA 5′- and 3′ ends to the respective intronic 5′- and 3′ ends, respectively.











(legend on next page)

RNAs within one family. To address this question, we compared iCLIP read counts for individual transcripts between relevant libraries (Figure 3A). This analysis revealed that all of the bound mRNAs (conventional and RDH RNAs), lncRNAs, and sn(o)RNAs displayed largely similar binding profiles for CBP20, ARS2, PHAX, and ZC3H18. To try to identify differently bound RNAs, we focused on PHAX and ZC3H18, which appeared to have the most diverse sets of targets (see Figure 1B). We performed a differential expression sequencing (DE-seq) analysis of their respective iCLIP reads, which demonstrated that of a total of 11,514 RNAs, 79% were bound indistinguishably by the two proteins, while 7% and 14% were bound preferentially by ZC3H18 and PHAX, respectively (Figure 3B). Most of the specific PHAX binding events occurred on snRNAs, in agreement with previous analyses (Figure 1B). We then focused on mRNAs and found that 74% of these targets were shared (Figure 3C). Taken together, these analyses thus indicate that even within single RNA families, CBP20, ARS2, PHAX, and ZC3H18 bind similar RNAs. This apparent lack of specificity was further confirmed by an analysis of the motifs enriched in the iCLIP reads: in agreement with binding primarily determined by cap proximity, no motifs were clearly identified other than CpG-rich sequences, which are generally enriched near TSSs, and U-rich sequences, which are prone to cross-linking (Figure S4). In addition, the enrichment scores for all pentameric motifs around the cross-linking sites were highly correlated for the different proteins (Figure S4). RBM7 generally showed the weakest correlation, in agreement with its more widespread binding to cap-distal regions.

We next analyzed whether transcripts of different lengths would reveal any differential binding. To this end, all analyzed capped RNAs were ranked by their length and the cumulative distribution of reads was computed (Figure 3D, left panel). This demonstrated a preference of PHAX and ARS2 for short RNAs, while RBM7 bound preferentially longer transcripts in agreement with its enrichment on pre-mRNAs. We then tested whether this effect was driven by all RNA families and therefore repeated the calculation after removal of snRNAs (Figure 3D, middle panel), or both snRNAs and histone mRNAs (Figure 3D, right panel). This demonstrated that these two families were largely responsible for the preferential binding of PHAX to small RNAs, leaving only limited size discrimination for the remaining transcripts.

Altogether, we conclude that CBP20, ARS2, PHAX, and ZC3H18 bind similar transcripts at steady state. For the large number of included mRNAs, we failed to detect any strong dependence on length for PHAX binding.

Steady-State RNA Binding of PHAX and ZC3H18 Correlates Poorly with Function

The surprise that PHAX and ZC3H18 bind similar RNAs despite having differently reported targets led us to ask whether the steady-state binding of these proteins correlated with transcript change upon factor depletion. Hence, we depleted PHAX or ZC3H18 by RNAi in HeLa cells and profiled the resulting mRNA contents by RNA-seq (Figure 3E). A DE-seq analysis against a control siRNA revealed that 422 mRNAs were significantly affected by ZC3H18 depletion, while none were significantly affected by PHAX depletion, despite similar depletion efficiencies (Log2 ratios of −2.4 and −1.7 for ZC3H18 and PHAX, respectively). This lack of effect of PHAX depletion on mRNAs was consistent with its known function as a pre-snRNA export factor but not with its iCLIP RNA binding profile, which displays robust mRNA binding.

We then considered separately the mRNAs that were preferentially bound by PHAX or by ZC3H18 (see Figure 3C). However, a similar fraction of mRNA was sensitive to the depletion of ZC3H18 regardless of its binding preference (Figure 3E), and a similar percentage of mRNA sensitive to ZC3H18 depletion was also identified in the entire mRNA population (Figure 3E). We conclude that the steady-state RNA binding profiles of PHAX and ZC3H18 correlate poorly with protein function at the genome-wide level.

ARS2 and ZC3H18 Link the CBC to NEXT

A way to rationalize that the interrogated factors largely share RNA targets, yet have a different effect, would be that these proteins are part of the same complex. However, while previous analyses showed that the CBCA complex can interact with PHAX (forming CBCAP; Hallais et al., 2013), and with ZC3H18 and NEXT (forming CBCN; Andersen et al., 2013), no interactions have yet been reported between PHAX and ZC3H18/NEXT.

Thus, to clarify these physical links further, we first determined protein-protein interactions between factors by performing pairwise two-hybrid assays of the human proteins in yeast cells (Y2H). As expected, robust interactions were detected between RBM7 and ZCCHC8 as well as between ZC3H18 and ARS2 (Table S3). Interactions of the CBC were monitored by co-expressing untagged CBP20 with CBP80 fused to the GAL4 DNA binding domain, together with the various preys fused to the GAL4 activation domain (Hallais et al., 2013). Using this strategy, we detected the expected interactions of the CBC with ARS2, PHAX, and NELF-E, a protein previously shown to directly interact with the CBC and used as a positive control (Narita

Figure 3. ARS2, PHAX, and ZC3H18 Are Targeted to Common Transcripts

(A) Scatterplots showing RPKM values of iCLIP tags from one indicated library versus another. Each RNA species is a dot. Gray, pre-mRNAs; violet, histone mRNAs; light blue, lncRNAs; and red, sn(o)RNAs.

(B) Scatterplot showing the log2 fold changes in PHAX versus ZC3H18 binding, as a function of normalized read counts for all RNAs identified in the iCLIP experiments. RNAs binding similarly to PHAX and ZC3H18 (gray dots) or significantly more to one protein (red dots) were determined by the DE-seq package.

(C) Venn diagram displaying mRNAs bound by PHAX (yellow) and/or ZC3H18 (green), as determined by DE-seq analysis of the iCLIP data.

(D) Cumulative distribution of iCLIP reads from the indicated replicate libraries ranked as a function of RNA size (x axis). Left: all capped RNAs; middle: all capped RNAs except snRNAs; and right: all capped RNAs except snRNAs and histone mRNAs.

(E) Bar plots displaying fractions of mRNA affected by ZC3H18 depletion (red) in the entire mRNA population (left) or in the mRNAs preferentially bound by PHAX or ZC3H18 (middle and left, respectively). For the same RNA population, the mean change in expression levels upon depletion of ZC3H18 is shown in blue. PHAX- and ZC3H18-bound mRNAs are shown in Figure 3C. The differences between the three populations are not statistically significant.

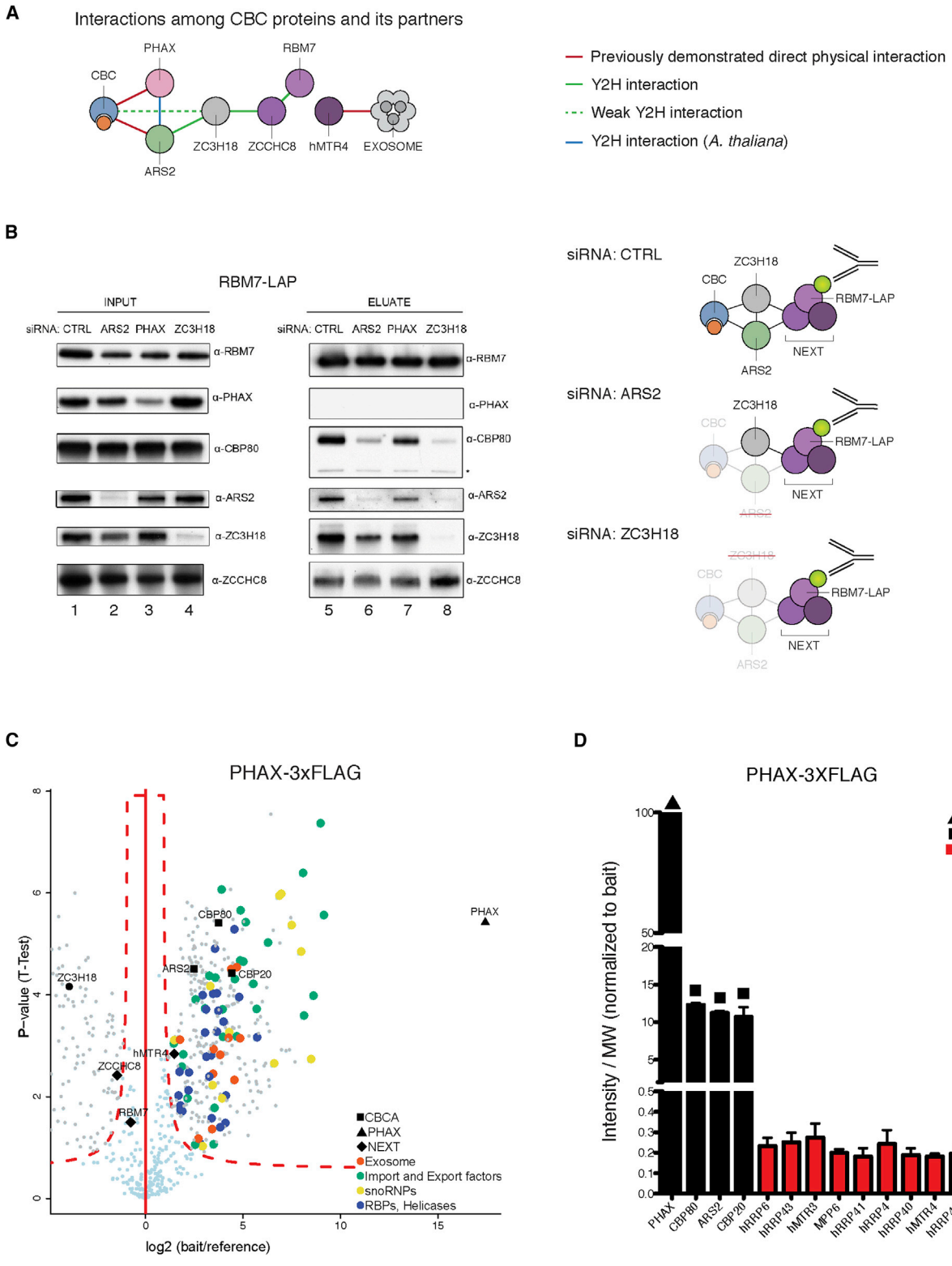


Figure 4. Molecular Organization of CBC-Related Complexes

(A) Schematic overview of Y2H data acquired from pairwise tests and cDNA library screens (see Table S3). The interaction of hMTR4 and the core exosome with RBM7/ZCCHC8 is indicated. The previously demonstrated direct physical interaction is from Andersen et al. (2013); Hallais et al. (2013); Lubas et al. (2011), and Ohno et al. (2000).

(legend continued on next page)

et al., 2007). Interestingly, a weak interaction was also detectable between the CBC and ZC3H18 (Table S3). To gather more data, we used human, *Drosophila*, and *Arabidopsis* ARS2 as well as human ZC3H18 as baits and performed Y2H screens of cDNA libraries of matched species. This recapitulated the ARS2-ZC3H18 interaction with *Drosophila* factors and revealed two interactions: (1) between the *Arabidopsis* homologs of ARS2 and PHAX and (2) between human ZC3H18 and ZCCHC8. The latter result was supported by the identification of a fragment located at the end of ZC3H18 (amino acids 746–953), which was sufficient to confer a robust interaction with ZCCHC8 in Y2H assays and co-IP experiments (Table S3; Figure S5A). The detected links of ARS2/ZC3H18 to the CBC and of ZC3H18 to the NEXT component ZCCHC8 suggested a collective interpretation of the Y2H results as depicted in Figure 4A. Consistent with previous affinity capture (AC)/mass spectrometry (MS) and in vitro protein-protein interaction data (Andersen et al., 2013; Hallais et al., 2013; Lubas et al., 2011), the CBC and NEXT complexes constitute separate entities with no apparent direct interaction. Instead, contact between CBC and NEXT appears to be mediated by ZC3H18 and ARS2. Moreover, PHAX, like ZC3H18, is capable of interacting with the CBC and ARS2 (Figure 4A; Hallais et al., 2013).

To substantiate the Y2H interaction results, we conducted a RBM7-LAP co-IP experiment and interrogated the ability of this NEXT component to associate with CBC-related factors in the presence or absence of ARS2, PHAX, or ZC3H18. Western blotting analysis of input samples from HeLa RBM7-LAP cells revealed that these three components were downregulated by administration of specific siRNAs, relative to control (CTRL) siRNAs (Figure 4B, lanes 1–4). RBM7 efficiently co-IPed CBP80, ARS2, ZC3H18, and the NEXT component ZCCHC8, whereas PHAX was undetectable (Figure 4B, lane 5). Consistently, depletion of PHAX did not change the RBM7 interaction pattern (Figure 4B, lane 7). In contrast, depletion of either ARS2 or ZC3H18 significantly decreased RBM7's interaction with CBP80 (Figure 4B, compare lanes 5, 6, and 8). Moreover, the ARS2-RBM7 association was lost upon ZC3H18 depletion and the contact between RBM7 and ZC3H18 was moderately affected by ARS2 depletion. None of the RNAi experiments affected the ability of the RBM7-LAP fusion to be captured by bead-bound GFP antibodies or its precipitation of the NEXT partner ZCCHC8. These results support the protein interactions suggested by the Y2H data and position ARS2 and ZC3H18 as critical factors bridging the CBC with the NEXT complex (Figure 4B, right panel).

The inability of RBM7 to IP PHAX (Figure 4B), and the absence of PHAX in IPs of NEXT components and ZC3H18 (Andersen et al., 2013), suggested that the majority of cellular NEXT/ZC3H18 and PHAX might reside in separate protein assemblies. Consistent with this notion, a PHAX-3xFLAG AC/MS experiment efficiently detected ARS2, CBP80, and CBP20 but failed to detect ZC3H18, ZCCHC8, and RBM7 (Figure 4C; Table S4). Human MTR4 was detected in low, yet significant, yields, which likely reflects its interaction with the exosome, the core subunits of which were detected at similar quantities (Figure 4D).

PHAX and ZC3H18 Compete for the CBC

Given their mutual exclusive presence in IP eluates, we considered that PHAX and ZC3H18 might compete for binding to the CBC. To investigate this possibility, RBM7-LAP interacting proteins were immobilized on GFP antibody-conjugated beads and challenged by increasing amounts of recombinant human PHAX produced in *E. coli*. In vitro, this recombinant protein was able to form a stable complex with the CBC (Figure S5B). In CTRL experiments without addition of exogenous protein or with 40 μg of added BSA, RBM7-LAP was retained on beads with CBP20, CBP80, ARS2, ZC3H18, and hMTR4 (Figure 5A, left panel lanes 4 and 6). In contrast, addition of PHAX caused CBP20, CBP80, and ARS2 to be dissociated in a concentration-dependent manner, whereas ZC3H18 and hMTR4 remained bead bound with RBM7-LAP (Figure 5A, left panel lanes 5–12). Thus, exogenous PHAX was capable of breaking the link between ZC3H18/NEXT and the CBC (Figure 5A, right panel), suggesting a competition between PHAX and ZC3H18 for binding the CBC.

Further support for this idea was obtained by employing the LUMIER assay, which yields a quantitative measure of the in vivo interaction between two proteins of interest (Figure 5B, left panel). A construct harboring CBP20 fused at its N terminus to the firefly luciferase (FFL) protein and 3xFLAG (3xFLAG-FFL-CBP20) was transfected into HEK293T cells together with a construct expressing either PHAX (RL-PHAX) or ZC3H18 (RL-ZC3H18) N-terminally fused to the Renilla luciferase protein. Subsequently, whole cell extracts were subjected to anti-FLAG IPs and luciferase activities were measured in both the input extracts and their IP pellets. As a measure of interaction specificity, Renilla luciferase (RL) was first plotted as fold enrichment over CTRL beads with no FLAG antibody, confirming that both RL-PHAX and RL-ZC3H18 exhibited robust interaction with 3xFLAG-FFL-CBP20 (Figure 5B, right panel). These interactions were then challenged by overexpression of putative

(B) Left: western blotting analysis of RBM7-LAP co-IP experiments conducted from extracts of HeLa cells depleted of factors using siRNAs as indicated. CTRL denotes CTRL siRNA targeting FFL mRNA. Input samples used for IP are shown to the left (lanes 1–4) and eluate samples from the IP are shown to the right (lanes 5–8). Right: schematics depict the interpretation of the conducted co-IPs.

(C) Volcano plot displaying the result of triplicate PHAX-3XFLAG AC/MS experiments. The log2 fold change of peptide MS intensities between bait and reference ("bait-less" cell line) eluate samples (x axis) were plotted against the negative log10 p values (y axis) calculated across the triplicate data (Student's t test). A dashed red curve separates specific PHAX-interacting proteins (upper right part of plot) from enriched proteins from the reference cell line (upper left part of plot). Some PHAX-interacting protein groups are color coded as indicated in the legend, and protein names relevant for this study are denoted. The full dataset of specific co-precipitants is given in Table S4.

(D) Column chart displaying abundance of selected proteins from PHAX-3XFLAG AC eluates. Peptide intensities divided by protein molecular weight (MW) were normalized to results for the bait protein. In this analysis, reference values were not subtracted from bait values, as the reference procedure yielded more background material binding to unshielded antibody epitopes sometimes obscuring analysis (data not shown). Note disruptions of the y axis to reveal intensities of all plotted factors.

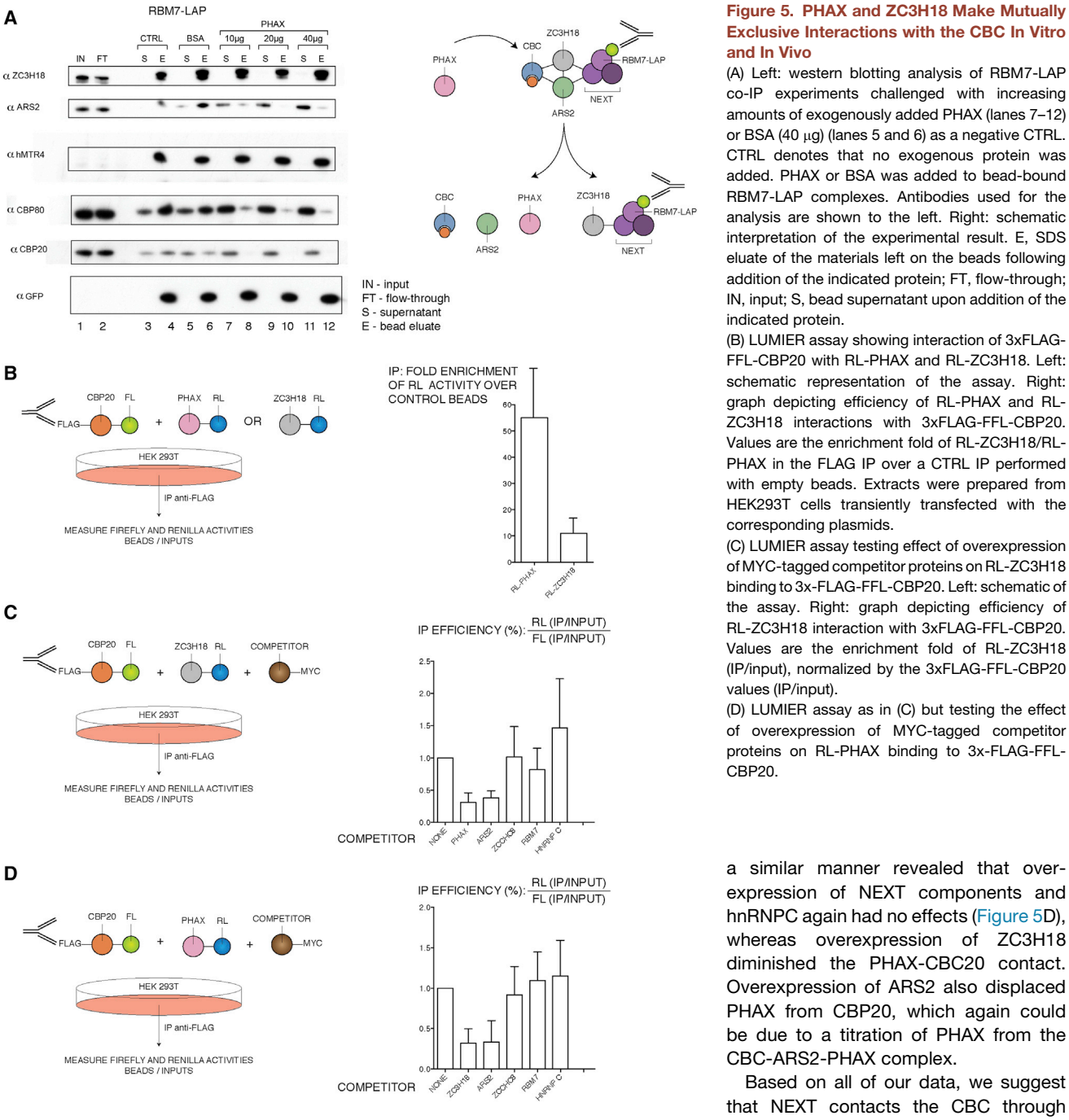


Figure 5. PHAX and ZC3H18 Make Mutually Exclusive Interactions with the CBC In Vitro and In Vivo

(A) Left: western blotting analysis of RBM7-LAP co-IP experiments challenged with increasing amounts of exogenously added PHAX (lanes 7–12) or BSA (40 μg) (lanes 5 and 6) as a negative CTRL. CTRL denotes that no exogenous protein was added. PHAX or BSA was added to bead-bound RBM7-LAP complexes. Antibodies used for the analysis are shown to the left. Right: schematic interpretation of the experimental result. E, SDS eluate of the materials left on the beads following addition of the indicated protein; FT, flow-through; IN, input; S, bead supernatant upon addition of the indicated protein.

(B) LUMIER assay showing interaction of 3xFLAG-FFL-CBP20 with RL-PHAX and RL-ZC3H18. Left: schematic representation of the assay. Right: graph depicting efficiency of RL-PHAX and RL-ZC3H18 interactions with 3xFLAG-FFL-CBP20. Values are the enrichment fold of RL-ZC3H18/RL-PHAX in the FLAG IP over a CTRL IP performed with empty beads. Extracts were prepared from HEK293T cells transiently transfected with the corresponding plasmids.

(C) LUMIER assay testing effect of overexpression of MYC-tagged competitor proteins on RL-ZC3H18 binding to 3x-FLAG-FFL-CBP20. Left: schematic of the assay. Right: graph depicting efficiency of RL-ZC3H18 interaction with 3xFLAG-FFL-CBP20. Values are the enrichment fold of RL-ZC3H18 (IP/input), normalized by the 3xFLAG-FFL-CBP20 values (IP/input).

(D) LUMIER assay as in (C) but testing the effect of overexpression of MYC-tagged competitor proteins on RL-PHAX binding to 3x-FLAG-FFL-CBP20.

competitor proteins (Figure S5C). Consistent with the proposed CBCN architecture (Figure 4A), overexpression of NEXT components had no effect on the ZC3H18-CBP20 interaction (Figure 5C, right panel). A similar result was obtained employing hnRNPC, another proposed CBC binder (McCloskey et al., 2012). However, in agreement with the in vitro experiments of Figure 5A, overexpression of PHAX readily displaced ZC3H18 from CBP20. ARS2 overexpression also decreased the interaction, possibly by titrating ZC3H18 from a CBC/ARS2/ZC3H18 ternary assembly. Challenging the PHAX-CBP20 interaction in a similar manner revealed that overexpression of NEXT components and hnRNPC again had no effects (Figure 5D), whereas overexpression of ZC3H18 diminished the PHAX-CBC20 contact. Overexpression of ARS2 also displaced PHAX from CBP20, which again could be due to a titration of PHAX from the CBC-ARS2-PHAX complex.

Based on all of our data, we suggest that NEXT contacts the CBC through Z3CH18 and ARS2, and that the formation of CBC-ARS2-PHAX and CBC-ARS2-ZC3H18 is mutually exclusive.

PHAX and ZC3H18 Have Opposite Effects on RNA Levels

Whereas our CLIP data showed that ZC3H18 and PHAX associate with the same set of RNAs, our biochemical experiments demonstrated that these factors cannot simultaneously bind the CBCA complex. This suggests that an RNA bound by CBCA may transition between complexes containing either ZC3H18 or PHAX. If these proteins elicit different functional outcomes,

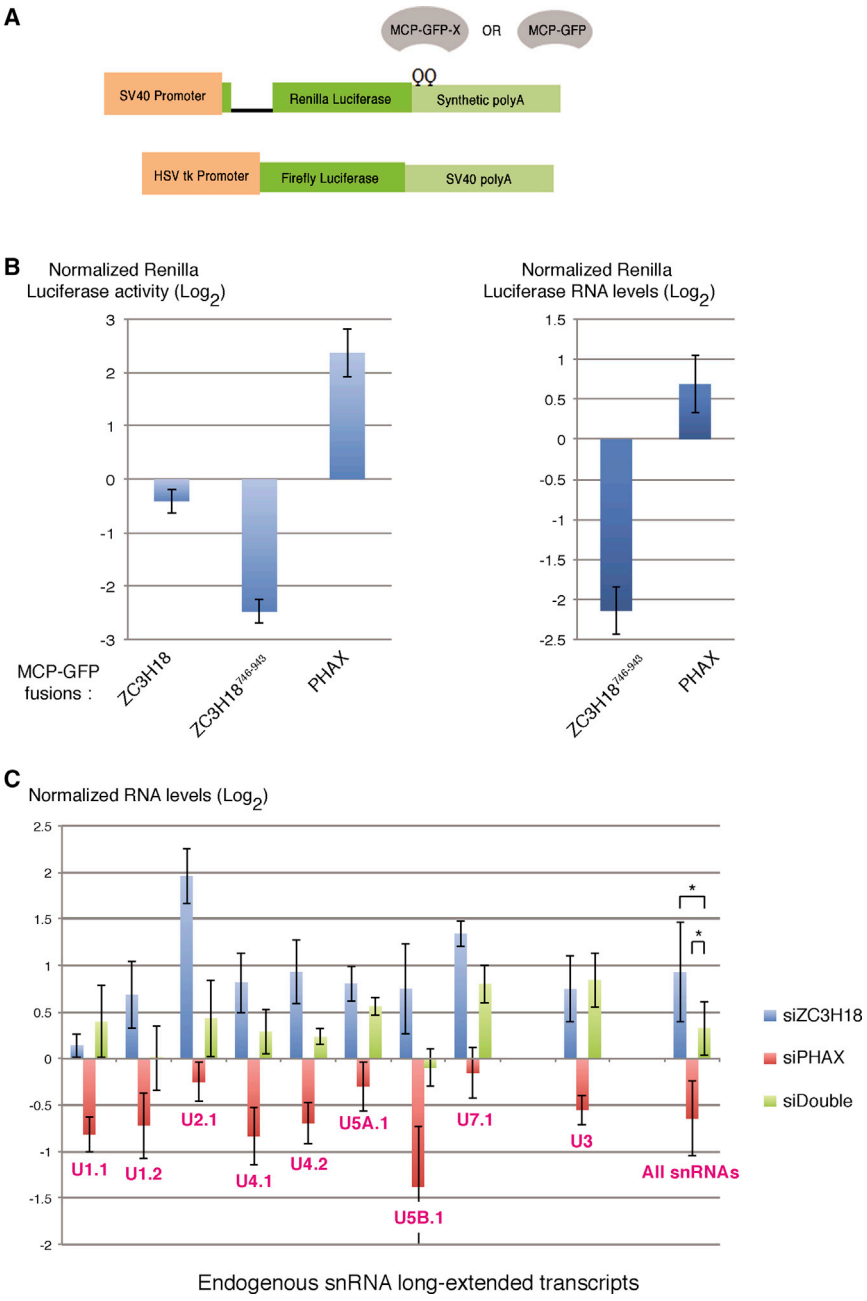


Figure 6. PHAX and ZC3H18 Exhibit Antagonistic Effects on RNA Levels

(A) Schematic representation of the employed tethering assay. An RL reporter RNA containing two MS2 binding sites in its 3′ UTRs was contained on a plasmid also harboring an FL reporter to CTRL for transfection efficiencies. This plasmid was co-transfected with a plasmid expressing candidate polypeptides fused to MS2-GFP (MCP-GFP-X) or with a plasmid expressing MS2-GFP alone.

(B) Effects on RL reporter activity of tethering MCP-GFP-X fusions. Left: RL/FFL activity ratios obtained with the MCP-GFP-X fusion and normalized to the same ratio derived from the corresponding MCP-GFP CTRL sample. Right: RL/FFL RNA ratios measured by qRT-PCR and expressed as Log_2 fold ratios between the MCP-GFP-X protein and the CTRL MCP-GFP fusion. Bars represent SDs from > 5 experiments.

(C) Effects of PHAX and ZC3H18 single- and double-depletions on levels of snRNA species carrying a long 3′ extension. Levels of the indicated transcripts were measured by qRT-PCR on RNA extracted from HeLa cells treated with the indicated siRNAs (color coded as displayed on the right). Values are displayed as Log_2 fold changes relative to samples treated with a CTRL FFL siRNA. Bars represent SDs from > 3 independent transfection experiments. Stars indicate significantly different values ($p < 0.02$ with a Student's t test).

RNA fate might then be dictated by which RNP complex is favored at the time this "decision" has to be made. To address the validity of this hypothesis, we first employed a tethering assay to explore the functional consequences of binding PHAX or ZC3H18 to an RNA reporter. Hence, we fused ZC3H18 or PHAX to the MS2 coat protein (MCP), which itself was fused to GFP (MCP-GFP-X), and we co-expressed one of these fusion proteins together with a plasmid expressing an RL RNA reporter carrying two MS2 binding sites in its 3′-UTR as well as a FFL CTRL RNA to adjust for transfection efficiencies (Figure 6A). Tethering of ZC3H18 decreased RL expression, which was likely due to recruitment of the NEXT complex, since tethering of the ZC3H18$^{746–953}$ fragment, sufficient for ZCCHC8 interaction (Table S3; Figure S5A), had a similar effect (Figure 6B, left panel). In stark contrast, tethering of PHAX induced a robust increase in RL activity. These effects were also reflected at the level of RL mRNA (Figure 6B, right panel).

To test the effects of PHAX and ZC3H18 on endogenous RNAs, we turned to snRNAs, whose long 3′ extended species are known to be degraded by the exosome in an ZC3H18/NEXT-dependent manner (Andersen et al., 2013), providing useful model substrates. As expected, depleting ZC3H18 generally increased levels of 3′-extended RNAs derived from eight different snRNA genes and the capped U3 snoRNA gene (Figure 6C; see depletion efficacy in Figure S6). In contrast, levels of the same substrates generally decreased upon PHAX depletion, whereas co-depletion of PHAX and ZC3H18 cancelled the effects of the individual depletions, which was also evident when averaging all snRNA substrates (Figure 6C, "all snRNAs"). Interestingly, the effect of co-depletion was not always simply the addition of the individual depletion effects. For instance, depletion of ZC3H18 had little effect on U1.1 3′-extended transcripts. However, it completely cancelled the negative effect of depleting PHAX, suggesting that ZC3H18 had gained access to these

RNAs in the absence of PHAX. Thus, the absence of one protein sensitized transcripts to the presence of the other. This is in line with a model where ZC3H18 and PHAX compete for RNA bound by CBCA to yield opposite functional outcomes.

PHAX and ZC3H18 Exchange Rapidly on the CBC In Vivo

The idea that CBCA-bound RNPs might transition between CBCA-PHAX and CBCA-ZC3H18 forms implies that PHAX and ZC3H18 do not simply bind and "mark" RNPs for different destinies. It also implies that PHAX and ZC3H18 rapidly exchange on and off the CBC. To test this prediction, we employed a LacO/Laci co-recruitment assay (Hallais et al., 2013) to measure the lifetime of these interactions in living U2OS cells. We tethered CBP20 to an array of genomic LacO sites, by fusing it to a red fluorescent version of the Laci protein (mRFP-Laci-CBP20). Transfected cells displayed a diffuse nuclear mRFP-Laci-CBP20 signal in addition to a concentrated bright spot, corresponding to the location of the LacO array (Hallais et al., 2013; Figure S7A). We next tested whether the mRFP-Laci-CBP20 "spot" would recruit its various partners. Indeed, co-transfected GFP-tagged versions of CBP80, ARS2, PHAX, and ZC3H18 concentrated in mRFP-Laci-CBP20 spots (Figure S7A, left and right panels). This recruitment was specific, as the proteins were not enriched in a CTRL spot formed by mRFP-Laci-KPNA3 (Figure S7B). We could also demonstrate that ARS2, PHAX, and ZC3H18 interactions were dependent on RNA, as a mutant form of CBP20 that does not bind the cap (F83A F85A; Mazza et al., 2002) failed to recruit these proteins, and yet did not prevent CBP80 interaction as expected (Figure S7C). In agreement with these results, we detected poly(A)$^+$ RNA accumulating in the mRFP-Laci-CBP20 spot (Figure S8), indicating that the tethered CBC binds capped RNAs. Our proteomic, LUMIER, and in vitro experiments showed that the interactions of the CBC with ARS2/PHAX/ZC3H18 are RNA independent (Figures 4 and 5; Hallais et al., 2013; Andersen et al., 2013). However, the CBC undergoes a large conformational change upon cap binding (Mazza et al., 2002). It is likely that this structural change is required for the CBC to bind its partners, thereby explaining its cap-dependent/RNA-independent interactions. Taken together, these data indicate that these CBC complexes are unlikely to bind nascent RNAs as a preformed species.

Having established a functional experimental design, we employed fluorescent recovery after photobleaching (FRAP) to measure the dynamics of mRFP-Laci-CBP20 interactions with its GFP-tagged partners. After photobleaching the LacO spot, the mRFP-Laci-CBP20 fluorescence showed very slow recovery over a 2-min time course, indicating stable binding of the fusion protein to the LacO array (Figure 7A, right panel). GFP-CBP80 recovered quickly when photobleached in the nucleoplasm, but only slowly (within minutes) in the mRFP-Laci-CBP20 spot, consistent with a stable interaction between these CBC subunits in vivo. In contrast, ARS2 and PHAX recovered quickly when photobleached in the Laci-CBP20 spot, with half-times of recovery of only a few seconds (Figures 7B and 7C). However, these kinetics were slower than recovery in the nucleoplasm, suggesting that dissociation of ARS2 and PHAX from the CBC is slower than the time it takes these molecules to diffuse through the bleached spot. Because ZC3H18 interacted with itself in the co-recruitment assay (Figure S7D), we performed the FRAP assay by tethering mRFP-Laci-ZC3H18 to the LacO array. This ensured that the photobleaching of GFP-CBP80 only measured the interaction between this protein and tethered ZC3H18 (see the Experimental Procedures). This revealed a rapid (within seconds) recovery of the GFP-CBP80 signal to the spot formed by mRFP-Laci-ZC3H18 (Figure 7D).

Modeling of the FRAP data showed that the lifetime of the CBP20-CBP80 interaction was in the order of minutes, whereas the lifetime of CBP20 interactions with ARS2, PHAX, or ZC3H18 was much shorter and in the range of 3–13 s (Table S5).

DISCUSSION

Eukaryotic cells produce various types of RNA that each follow a certain processing/decay and/or transport pathway. How proper transcript sorting into appropriate pathways occurs is a fundamental but incompletely understood problem. Because the CBC promotes the processing of different RNAs, yielding family-specific effects (Gonatopoulos-Pournatzis and Cowling, 2014; Müller-McNicoll and Neugebauer 2014), it provides an interesting model to study the concept of RNA sorting. It has been suggested that such family- or transcript-specificity derives from CBC partners binding only certain RNAs, hereby acting as identity marks (Ohno et al., 2002). Our results do not support this idea, but instead suggest an alternative model where early RNP complexes are constantly remodeled and determine RNA fate by reacting to external input at specific times during RNA biogenesis.

Binding of Some Landmark RNA Binding Proteins Is Promiscuous and Not Sufficient to Define RNA Maturation Pathways

Early studies in *Xenopus* oocytes demonstrated that distinct RNA families use non-overlapping nuclear export pathways (Jarmolowski et al., 1994). Consistently, it was found that pre-snRNAs and mRNAs use distinct exportins and export adaptors: PHAX/CRM1 for pre-snRNAs (Ohno et al., 2000), and TAP, in association with ALYREF or other RNA binding proteins (RBPs), for mRNAs (Björk and Wieslander, 2014; Segref et al., 1997). Such specificity for a given export pathway appeared to stem from specific binding of key RBPs, such as PHAX or the EJC, to pre-snRNAs and spliced mRNAs, respectively (Ohno et al., 2002). This further suggested the possibility that RNA identity could be determined early on in the nucleus, perhaps even during transcription, and then stably maintained due to specific RNA coating by certain RBPs. The iCLIP data presented here do not support this hypothesis. This is because we detect binding of PHAX not only to pre-snRNAs as expected, but also to a large range of other capped RNAs, including PROMPTs, eRNAs, lincRNAs, RDH RNAs, and polyadenylated mRNAs. In fact, the fraction of total PHAX iCLIP reads mapping to mRNA approaches 40%, and is not restricted to particular mRNA species, not even to short transcripts as would perhaps have been predicted. When compared to CBP20, which expectedly binds to all capped RNAs, PHAX exhibits some preference for pre-snRNAs, but this specificity is moderate. With the notable exception of intronic snoRNAs, it is also important to note that binding of PHAX to RNA is likely to occur mainly through the CBC, which

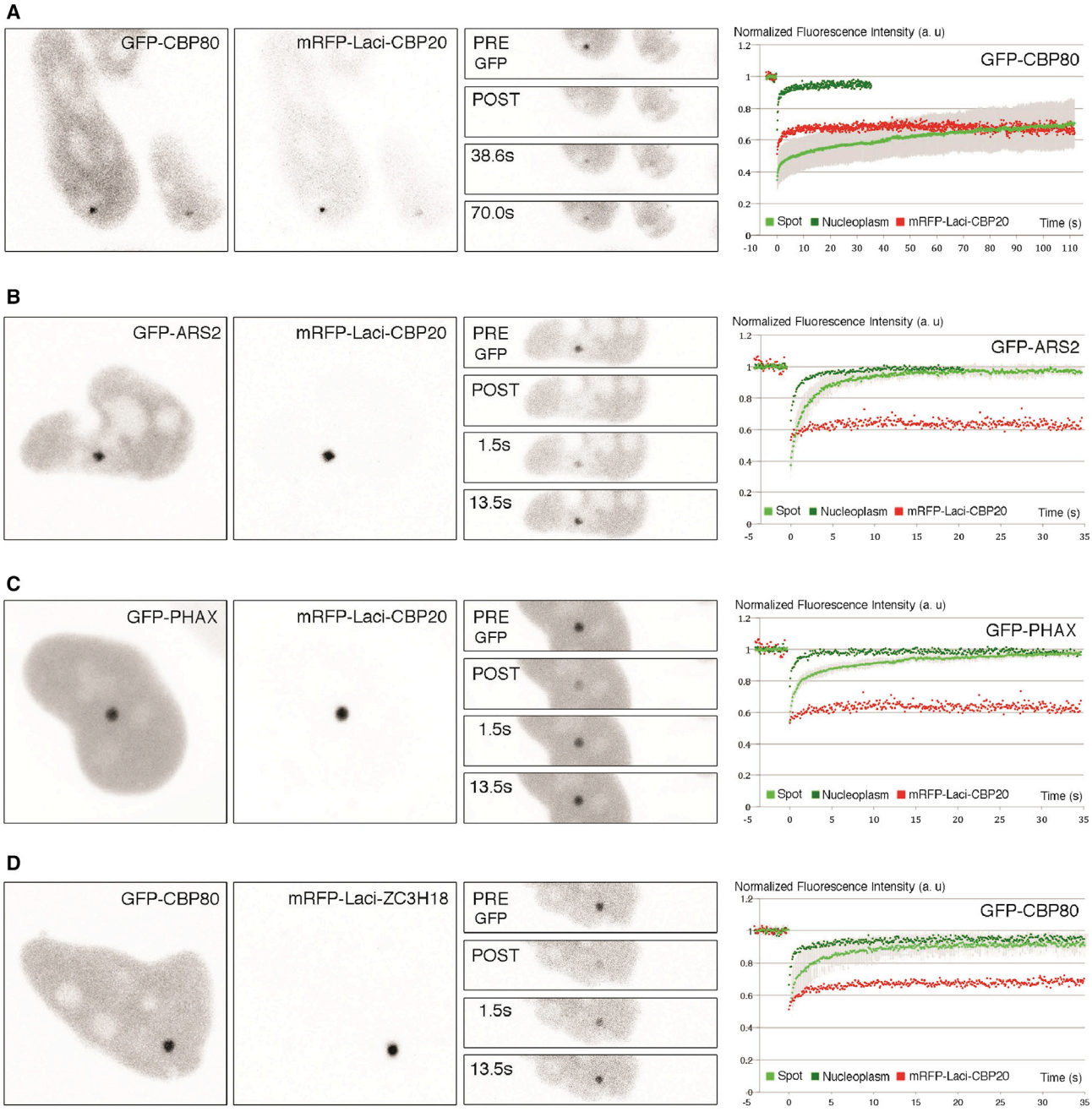


Figure 7. PHAX, ARS2, and ZC3H18 Exchange Rapidly on the CBC In Vivo

(A–C) Left: confocal images of U2OS cells carrying a LacO array and co-transfected with plasmids expressing the indicated proteins (fields of view are 30 × 30 μm; left, GFP; right, mRFP). Middle: confocal images of a FRAP experiment of the GFP-tagged protein (fields of view are 10 × 48 μm). Right: fluorescent recovery curves of the indicated proteins. The FRAP experiments in the green and red channels were performed independently. Dark green: the indicated GFP-tagged protein in the nucleoplasm; light green: the indicated tagged protein in the LacO spot; and red: the mRFP-Laci-CBP20 fusion in the LacO spot. y-axes denote fluorescence intensities corrected for photobleaching and normalized to pre-bleach intensities. x-axes denote time in seconds. Gray bars represent SDs calculated from > 10 different cells. (A: GFP-CBP80; B: GFP-ARS2; C: GFP-PHAX).

(D) As in (A) to (C), except that ZC3H18 was fused to Laci and tethered to the LacO spot in place of CBP20.

can be appreciated by the largely cap-proximal binding of the protein (see Figures 1E and 2). The limited target specificity of PHAX is thus probably not due to promiscuous RNA binding, but rather to its loading onto RNA via cap-bound CBC. Binding of even a key RBP like PHAX is therefore poorly discriminating. It may even be argued that PHAX is a bona fide mRNA binding protein and that it could have a previously unnoticed role in mRNA biogenesis. However, PHAX depletion revealed little

effect on steady-state mRNA levels or splicing patterns in transcriptome-wide experiments. Furthermore, steady-state binding of PHAX and ZC3H18, as determined by iCLIP, correlated poorly with effects on RNA levels upon depletion of these proteins (see Figure 3E). Using PHAX and ZC3H18 as a paradigm, we therefore suggest that binding specificity per se may generally not be sufficient to identify RNAs and determine their fate. A notable exception may be the EJC, which binds stably to spliced RNA and thus provides a more definitive identity mark (Le Hir et al., 2000a, 2000b). However, the EJC is deposited as a result of splicing, and it is thus a stable label for a transient phenomenon, much like the poly(A) tail is for 3′ end processing.

Mutually Exclusive Formation of CBC Complexes at Specific Maturation Checkpoints May Determine RNA Fate

Live cell imaging of RBPs has demonstrated their transient interaction with RNA, allowing rapid sampling of sequences. In agreement, our FRAP data show that CBC-containing complexes are quite labile, with a half-life of only a few seconds. With RNAPII elongation rates of about 2 kb/min (Boireau et al., 2007; Jonkers et al., 2014), a medium-sized human gene takes ~50 min to transcribe. Splicing and mRNA export also takes minutes (Audibert et al., 2002; Beyer and Osheim, 1988; Schmidt et al., 2011). This suggests that PHAX and ZC3H18 continuously exchange at the CBC-bound cap during RNA production. Thus, instead of using steady-state binding as a mechanism to identify RNAs and control their fate, many RBPs, including PHAX and ZC3H18, might be part of a "hit-and-run" mechanism, where transcript fate would originate from "locking" of decisive complexes only at particular checkpoints during pre-mRNA processing. The ability of RNPs to form mutually exclusive complexes with proteins having opposing activities may reflect the need of the RNP to keep all options open until one outcome would have to be selected out of several possibilities. Indeed, it may simply reflect the fact that RNAPII "does not know" which type of transcription unit it is engaged with until relevant cues are instigated.

We suggest that one such cue, or checkpoint, may occur when a 3′ end processing signal emerges from the RNAPII exit channel. Processing signals drive the assembly of specific proteins, which may then synergize with the CBC to lock the proper complex and produce the required outcome. In support of this model, CBCA was shown to stimulate the usage of a range of 3′-end processing signals (Hallais et al., 2013). Moreover, NEXT complex components purify with 3′-end processing factors (Shi et al., 2009). Thus, a cryptic, cap-proximal 3′-end/termination signal might promote an interaction between the CBCA complex at the RNA 5′ end with NEXT at the 3′ end, via ZC3H18. This would stabilize the CBCN complex, which would serve to exclude PHAX while simultaneously increase the access of NEXT and the exosome to the RNA 3′ end. Example substrates for such a scenario would be PROMPTs, whose early termination and degradation rely on promoter proximal poly(A) sites as well as the CBCA, NEXT, and exosome complexes (Andersen et al., 2013; Ntini et al., 2013). In contrast, the 3′-end processing signal of an snRNA would recruit the Integrator complex (Baillat et al., 2005), which might bias the competition between PHAX and ZC3H18 toward the formation of the CBCAP complex (Hallais et al., 2013), excluding ZC3H18/NEXT and resulting in productive 3′-end formation. If proper 3′-end formation is missed, such as in the case of "long 3′-extended" sn(o)RNAs, downstream cryptic termination sites might again favor CBCN formation and transcript decay.

In this study, we have focused on RNA transport via PHAX and RNA decay via ZC3H18/NEXT. However, because the CBC has many activities, it is likely that dynamic exchanges of mutually exclusive protein complexes at RNA caps may also interplay with other processing events, such as RNA splicing. We propose that the constant remodeling of CBC-associated complexes allows the dynamic integration of a diverse source of signals, whereas a pre-determined, rigid CBC complex, deposited for instance at the start of transcription, would not allow such regulation.

EXPERIMENTAL PROCEDURES

Cell Culture and Plasmids

HeLa, U20S, and HEK293 cells were grown in DMEM containing 10% fetal bovine serum (FBS) and 1% penicillin/streptavidin, at 37°C, 5% CO_2. DNA cloning was performed using standard techniques and the Gateway system (Invitrogen). The two-hybrid plasmids were based on pACTII, p422, and pAS2 dd (Hallais et al., 2013). Detailed maps and sequences are available upon requests.

siRNAs

Cells were transfected for 3 days using Lipofectamine 2000 (20 μL/mL in the transfection mixture, together with 0.4 μM siRNA), at a final siRNA concentration of 20 nM in the cell culture medium. siRNA sequences are indicated in the Supplemental Experimental Procedures.

iCLIP and Bioinformatic Analysis

The iCLIP approach was performed as described in Konig et al. (2011) with the additional modifications of Lubas et al. (2015), which include differences in sonication and washing buffers. iCLIP cDNA libraries were sequenced from two replicate experiments for each interrogated factor. Trimmed reads were mapped to the hg19 human genome assembly and genomic annotations were assigned based on gene annotations from the UCSC genome browser and published datasets. To compare the CLIP data with total RNA abundances, we used representative RNA-seq datasets downloaded from the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra). We used cytoplasmic poly(A)$^+$-selected data from HeLa (SRR3479116; Lykke-Andersen et al., 2014) and HEK293 (SRR1275413) cells, as well as rRNA-depleted total RNA from HeLa (SRR1014903) and HEK293 (SRR2096982) cells. RNA-seq data were analyzed with the same pipeline as iCLIP.

qRT-PCR Assays

For qRT-PCR analysis, RNAs were treated with DNase RQ1 (Sigma) for 1 hr at 37°C to digest residual genomic DNA. RT and qPCR were performed as indicated in the Supplemental Experimental Procedures.

Protein Interaction Assays

For AC/MS analysis, we used HEK293 Flp-In T-Rex cells stably expressing C-terminally 3xFLAG-tagged PHAX under CTRL of a tetracycline-inducible promoter. Cryogenic disruption of cells and 3xFLAG-AC methodology were performed as previously described (Andersen et al., 2013). For the PHAX competition assay, CBCN assembly was first immobilized on the magnetic beads by co-IP of RBM7-LAP (as above) and then challenged with recombinant PHAX or BSA as the CTRL. Proteins were analyzed by western blotting.

Yeast two-hybrid assays were performed as previously described (Hallais et al., 2013). Strains expressing preys and baits were crossed and diploids were plated on triple and quadruple selective media (−Leu/−Trp/−Ade

or −Leu/−Trp/−Ade/−His). Growth was assessed visually after 3 days at 30°C. A similar protocol was used for regular two-hybrid assays, except that p422 plasmids and adenine selection were omitted.

For LUMIER assays, cells were extracted in HNTG 2 days after transfection, and antibody-coated beads were incubated with extracts for 2 hr at 4°C. Beads were washed three times in HNTG and resuspended in passive lysis buffer (PBL) (Promega), and luciferase activity was measured in the inputs and pellets using the dual-luciferase assay (Promega). HNTG is 20 mM HEPES, pH 7.9, 150 mM NaCl, 1% Triton, 10% glycerol, 1 mM $MgCl_2$, 1 mM EGTA, and protease inhibitors (Roche).

MS2 Tethering Assay

HEK293 cells were co-transfected with the luciferase reporter plasmid containing two MS2 stem-loops in its 3′ UTR and with plasmids expressing MCP-GFP fused to the protein of interest. Two days later, cells were lysed in PBL buffer (Promega) and firefly and RL activities were measured as described above.

Microscopy and LacO FRAP Assay

U2OS cells carrying a LacO array were plated on coverslips and co-transfected using JetPrime (PolyPlus) with plasmids expressing the GFP fusion of interest together with the mRFP-Laci fusion of interest. Two days later, cells were either fixed and visualized by wide-field microscopy or imaged live using a Zeiss LSM780 microscope. FRAP was performed on a spot with a radius of 1.5 μm using 10 iterations at full laser power, and images were collected every 96 ms. The mean fluorescence intensities of a bleached and of a non-bleached area were calculated for each time point (I_{spot} and I_{cell}). The background signal was measured outside the cell (I_{bkg}). The bleaching and background corrected fluorescence intensity was then calculated at each time point $I = (I_{spot} - I_{bkg})/(I_{cell} - I_{bkg})$. This value was then normalized to 1 by dividing it with the value of I computed with the averaged pre-bleach time points.

ACCESSION NUMBERS

The accession number for the raw data from RNA-seq experiments reported in this paper is GEO: GSE94427.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, eight figures, and five tables can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2017.02.046.

AUTHOR CONTRIBUTIONS

S.G. performed iCLIP experiments, with the help of M.L.; S.G. and G.K. analyzed the iCLIP data; N.E.B. performed the experiments of Figures 5B–5D and 6 and Table S3, with the help of N.M. and M.-C.R.; N.E.B. and E.B. performed the experiment of Figure 7; M.D. performed the experiments of Figures 4 and 5A, with the help of J.B. and J.S.A.; W.M.S. produced and characterized the recombinant human PHAX protein; C.V. validated the iCLIP data (Figure S2) and supervised N.E.B.; E.B., T.H.J., and G.K. conceived the project and wrote the manuscript.

ACKNOWLEDGMENTS

We thank I. Poser and A. Hyman for the gift of the BAC-LAP cell lines and E. Izaurralde for anti-CBP20 and anti-CBP80 antibodies. Work in the T.H.J. laboratory was supported by the European Research Council (ERC) (grant 339953), Danish National Research Council, Danish National Research Foundation (grant DNRF58), Lundbeck Foundation, and Novo Nordisk Foundation. S.G. was partly supported by an Eiffel PhD fellowship from the Eiffel Excellence Scholarship Programme. Work in the E.B. laboratory was supported by a grant from the Ligue Nationale Contre le Cancer. G.K. was supported by the Wellcome Trust (grant 097383) and the UK Medical Research Council. N.E.B was supported by a fellowship from the Algerian Ministry of Higher Education and the Ligue Nationale Contre le Cancer (GB/MA/VSP-11059).

Received: September 1, 2016
Revised: December 13, 2016
Accepted: February 14, 2017
Published: March 14, 2017

REFERENCES

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.

Audibert, A., Weil, D., and Dautry, F. (2002). In vivo kinetics of mRNA splicing and transport in mammalian cells. Mol. Cell. Biol. *22*, 6706–6718.

Baillat, D., Hakimi, M.A., Näär, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell *123*, 265–276.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. *15*, 163–175.

Beyer, A.L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. *2*, 754–765.

Björk, P., and Wieslander, L. (2014). Mechanisms of mRNA export. Semin. Cell Dev. Biol. *32*, 47–54.

Boireau, S., Maiuri, P., Basyuk, E., de la Mata, M., Knezevich, A., Pradet-Balade, B., Bäcker, V., Kornblihtt, A., Marcello, A., and Bertrand, E. (2007). The transcriptional cycle of HIV-1 in real-time and live cells. J. Cell Biol. *179*, 291–304.

Boulon, S., Verheggen, C., Jady, B.E., Girard, C., Pescia, C., Paul, C., Ospina, J.K., Kiss, T., Matera, A.G., Bordonné, R., and Bertrand, E. (2004). PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. Mol. Cell *16*, 777–787.

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.

Chen, Y., Pai, A.A., Herudek, J., Lubas, M., Meola, N., Järvelin, A.I., Andersson, R., Pelechano, V., Steinmetz, L.M., Jensen, T.H., and Sandelin, A. (2016). Principles for RNA metabolism and alternative transcription initiation within closely spaced promoters. Nat. Genet. *48*, 984–994.

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.

Flaherty, S.M., Fortes, P., Izaurralde, 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. USA *94*, 11893–11898.

Glover-Cutter, K., Kim, S., Espinosa, J., and Bentley, D.L. (2008). RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nat. Struct. Mol. Biol. *15*, 71–78.

Gonatopoulos-Pournatzis, T., and Cowling, V.H. (2014). Cap-binding complex (CBC). Biochem. J. *457*, 231–242.

Görnemann, J., Kotovic, K.M., Hujer, K., and Neugebauer, K.M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol. Cell *19*, 53–63.

Gruber, J.J., Olejniczak, S.H., Yong, J., La Rocca, G., Dreyfuss, G., and Thompson, C.B. (2012). Ars2 promotes proper replication-dependent histone mRNA 3′ end formation. Mol. Cell *45*, 87–98.

Hallais, M., Pontvianne, F., Andersen, P.R., Clerici, M., Lener, D., Benbahouche, Nel.H., 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.

Ideue, T., Sasaki, Y.T., Hagiwara, M., and Hirose, T. (2007). Introns play an essential role in splicing-dependent formation of the exon junction complex. Genes Dev. *21*, 1993–1998.

Izaurralde, E., Stepinski, J., Darzynkiewicz, E., and Mattaj, I.W. (1992). A cap binding protein that may mediate nuclear export of RNA polymerase II-transcribed RNAs. J. Cell Biol. *118*, 1287–1295.

Izaurralde, 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.

Jarmolowski, A., Boelens, W.C., Izaurralde, E., and Mattaj, I.W. (1994). Nuclear export of different classes of RNA is mediated by specific factors. J. Cell Biol. *124*, 627–635.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. eLife *3*, e02407.

Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2011). iCLIP–transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. J. Vis. Exp. *50*, 2638.

Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000a). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. EMBO J. *19*, 6860–6869.

Le Hir, H., Moore, M.J., and Maquat, L.E. (2000b). Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. Genes Dev. *14*, 1098–1108.

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.

Lubas, M., Andersen, P.R., Schein, A., Dziembowski, A., Kudla, G., and Jensen, T.H. (2015). The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. Cell Rep. *10*, 178–192.

Lykke-Andersen, S., Chen, Y., Ardal, B.R., Lilje, B., Waage, J., Sandelin, A., and Jensen, T.H. (2014). Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. Genes Dev. *28*, 2498–2517.

Masuyama, K., Taniguchi, I., Kataoka, N., and Ohno, M. (2004). RNA length defines RNA export pathway. Genes Dev. *18*, 2074–2085.

Mazza, C., Segref, A., Mattaj, I.W., and Cusack, S. (2002). Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. EMBO J. *21*, 5548–5557.

McCloskey, A., Taniguchi, I., Shinmyozu, K., and Ohno, M. (2012). hnRNP C tetramer measures RNA length to classify RNA polymerase II transcripts for export. Science *335*, 1643–1646.

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.

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.

Ntini, E., Järvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jørgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R., et al. (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. Nat. Struct. Mol. Biol. *20*, 923–928.

Ohno, M., Sakamoto, H., and Shimura, Y. (1987). Preferential excision of the 5′ proximal intron from mRNA precursors with two introns as mediated by the cap structure. Proc. Natl. Acad. Sci. USA *84*, 5187–5191.

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.

Ohno, M., Segref, A., Kuersten, S., and Mattaj, I.W. (2002). Identity elements used in export of mRNAs. Mol. Cell *9*, 659–671.

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·U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. RNA *19*, 1054–1063.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851–1854.

Schmidt, U., Basyuk, E., Robert, M.C., Yoshida, M., Villemin, J.P., Auboeuf, D., Aitken, S., and Bertrand, E. (2011). Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation. J. Cell Biol. *193*, 819–829.

Segref, A., Sharma, K., Doye, V., Hellwig, A., Huber, J., Lührmann, R., and Hurt, E. (1997). Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores. EMBO J. *16*, 3256–3271.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3′ processing complex. Mol. Cell *33*, 365–376.

Worch, R., Niedzwiecka, A., Stepinski, J., Mazza, C., Jankowska-Anyszka, M., Darzynkiewicz, E., Cusack, S., and Stolarski, R. (2005). Specificity of recognition of mRNA 5′ cap by human nuclear cap-binding complex. RNA *11*, 1355–1363.