



HAL
open science

CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing

Marie Hallais, Frédéric Pontvianne, Peter Refsing Andersen, Marcello Clerici, Daniela Lener, Nour El Houda Benbahouche, Thierry Gostan, Franck Vandermoere, Marie-Cécile Robert, Stephen Cusack, et al.

► **To cite this version:**

Marie Hallais, Frédéric Pontvianne, Peter Refsing Andersen, Marcello Clerici, Daniela Lener, et al.. CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. Nature Structural and Molecular Biology, 2013, 20 (12), pp.1358–66. 10.1038/nsmb.2720 . hal-02191615

HAL Id: hal-02191615

<https://hal.science/hal-02191615v1>

Submitted on 26 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CBC-ARS2 stimulate 3'-end maturation of multiple RNA families and favor cap-proximal processing

Marie Hallais¹, Frédéric Pontvianne¹, Peter Refsing Andersen², Marcello Clerici³, Daniela Lener^{1,4}, Nour El Houda Benbahouche¹, Thierry Gostan¹, Franck Vandermoere⁵, Marie-Cécile Robert¹, Stephen Cusack³, Céline Verheggen^{1,*}, Torben Heick Jensen² and Edouard Bertrand^{1,*}

Running Title: CBCA and 3'-end formation

Key words: RNA biogenesis, cap, CBC, 3'-end processing, snRNA, capped snoRNAs, telomerase RNA, mRNA

1-Equipe labellisée Ligue contre le Cancer; IGMM-CNRS UMR 5535; 1919 route de Mende; 34293 Montpellier Cedex 5; France.

2-Centre for mRNP Biogenesis and Metabolism, Department of Molecular Biology and Genetics, Aarhus University, C. F. Møllers Allé 3, Bldg. 1130, DK-8000 Aarhus C, Denmark

3-EMBL Grenoble Outstation, 6 Rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France.

4- Present address: IBMC - CNRS UPR 9002; 15 Rue René Descartes; 67084 Strasbourg Cedex; France.

5- CNRS , UMR-5203, Institut de Génomique Fonctionnelle, F-34000 Montpellier, France ;
Inserm , U661, F-34000 Montpellier, France ; Universités de Montpellier 1 & 2 , UMR-5203,
F-34000 Montpellier, France.

*: to whom correspondence should be addressed: edouard.bertrand@igmm.cnrs.fr;

celine.verheggen@igmm.cnrs.fr

Summary :

The nuclear cap-binding complex (CBC) stimulates multiple steps in several RNA maturation pathways, but how it functions is incompletely understood. For snRNAs, the CBC recruits the exportin CRM1 through the adaptor protein PHAX. Here, we identify the CBCAP complex, composed of CBC, Ars2 and PHAX, and we show that both CBCAP and CBC-Ars2 complexes can be reconstituted from recombinant proteins. Ars2 stimulates PHAX binding to the CBC as well as snRNAs 3'-end processing, hereby coupling snRNA maturation with export. In vivo, CBC and Ars2 bind a similar set of capped non-coding and coding RNAs, and stimulate 3' processing at polyadenylation (pA) sites. Interestingly, strongest effects are displayed for cap-proximal pA sites, thereby favoring premature transcription termination. Ars2 functions through the mRNA 3'-end cleavage factor CLP1, which binds elongating Pol-II through the CTD-binding protein PCF11. Ars2 is thus a major effector of the CBC that stimulates real and cryptic 3'-end processing signals.

Introduction

Production of mature, functional RNAs requires a large number of steps, the coupling of which has become an intense subject of investigation ¹⁻³. The m7G cap is a crucial element that is present on all RNA polymerase II transcripts (Pol-II ⁴). It is added co-transcriptionally after the synthesis of the first 20 nucleotides ⁵, by a complex of enzymes that directly binds the S5 phosphorylated C-terminal domain of Pol-II ^{6,7}. As such, it is the hallmark of successfully initiated transcripts. The cap protects RNAs from exonucleolytic degradation and has numerous other functions. In the nucleus, it is bound by the nuclear cap-binding complex (CBC), which consists of CBP80 and CBP20 ⁸. The CBC has been shown to promote splicing of cap-proximal introns ⁸, 3'-end formation ⁹, intra-nuclear transport ¹⁰, export ^{11,12} and translation ¹³. It is thus a central node in RNA metabolism, which potentially integrates and couples these steps.

Most interestingly, the CBC elicits distinct effects for different RNA families. For snRNAs, the CBC promotes nuclear export of their precursors through the adaptor PHAX, which in turns recruits the exportin CRM1 ¹². For polyadenylated mRNAs, the CBC also stimulates export, but in this case acts by bridging the cap to the export receptor TAP through the adaptor ALY ¹¹. The CBC further promotes mRNA splicing, 3'-end formation and it facilitates the pioneer round of translation by recruiting CTIF ^{8,9,13,14}. In the case of non-polyadenylated histone mRNAs, the CBC has also been shown to promote 3'-end formation, and this was shown to involve an interaction with NELF and SLBP ¹⁵. These observations raise several important questions. First, despite its importance in RNA metabolism, only PHAX, NELF, and CTIF have been formally shown to directly contact the CBC in humans, leaving some important effectors of the CBC yet to be discovered. In particular, the factors that mediate the effects of the CBC on splicing and on cleavage and polyadenylation still remain elusive. Second, although it is clear that the CBC has different effects depending on the type of RNA it binds, the basis for such discrimination of different classes of RNAs is not well understood. This is however critical

because the CBC must activate the proper machinery to correctly process the RNAs and to route them to the correct biogenesis pathways¹⁶. One important feature that discriminates families of capped RNAs is their length, with short RNAs being specifically associated with PHAX and long ones with ALY¹⁶. The third question relates to coupling. This occurs when one processing event stimulates the following one¹⁻³. Because the CBC stimulates several processing steps, it is ideally placed to couple them, but whether this occurs, and how, is not known.

A useful paradigm to study both the coupling and the specificity of CBC action is PHAX, because this factor appears to act specifically on small-sized RNAs like pre-snRNAs and capped pre-snoRNAs, and not on long RNAs like mRNAs^{10,16}. Here, we purified PHAX from nuclear extracts and identified the CBCAP complex, which also contains CBC and Ars2. We show that Ars2 couples snRNA 3'-end processing with export through the CBC. Surprisingly, CBC and Ars2 also stimulate 3'-processing of polyadenylated sites, and they preferentially act on cap-proximal sites. Ars2 thus appears to be a general adaptor that links the cap to 3'-end maturation, and that favor the production of short RNAs.

Results

The CBCAP complex contains CBC, Ars2 and PHAX.

To better understand PHAX function, we purified it from HeLa nuclear extracts and characterized its associated proteins by mass spectrometry. To this end, we generated highly specific monoclonal antibodies, which revealed a single band in western blots (Supplementary Figure 1A). Proteins immuno-precipitated from control beads or beads coated with anti-PHAX antibodies were separated by gel electrophoresis and stained with coomassie blue. Several bands were present in high and roughly stoichiometric amounts in the PHAX immuno-precipitates (IP; Figure 1A). Mass-spectrometry indicated that two of the major bands corresponded to CBP20 and CBP80 as expected, while the third one was a protein of an apparent molecular weight of 120 kD and identified as Ars2. The similar abundance of these proteins in the IP suggested that PHAX forms a stoichiometric complex with both CBC and Ars2.

Ars2 has been previously shown to co-purify with the CBC ¹⁷, and our results thus indicated that at least a fraction of this complex also contains PHAX. To investigate this possibility, we attempted to co-IP the two proteins. As expected, Ars2 was co-precipitated with anti-PHAX antibodies from both nuclear and cytoplasmic S100 extracts (Figure 1B), and this interaction was resistant to RNase treatment, indicating that it is due to protein-protein contacts rather than being mediated by RNA. Given that association of proteins can occur in extracts, after cell lysis, we performed live cell experiments using a co-recruitment assay ¹⁸. PHAX was fused to Laci-mCherry and co-expressed with GFP-Ars2 in U2OS cells containing a large LacO array inserted in their chromatin, and forming a large nucleoplasmic spot in interphasic cells (Figure 1C). PHAX-Laci-mCherry accumulated at the LacO array and recruited GFP-Ars2, while a non-interacting protein like GFP-KPNA2 remained diffusely distributed in the nucleoplasm. Likewise, KPNA2-Laci-mCherry failed to recruit GFP-Ars2 to the spot

(Supplementary Figure 1B), thus demonstrating a specific interaction between PHAX and Ars2 in live cells.

We then dissected the interactions taking place using two-hybrid tests. Surprisingly, no interaction could be detected between either individual CBC subunits and Ars2 or PHAX, or between PHAX and Ars2 (Supplementary Figure 1C). In fact, the only positive interaction was between CBP20 and CBP80. We reasoned that some of the interactions might occur only with the complete CBC complex, rather than with its individual subunits. To test this, we used a bridged two-hybrid approach, in which the two CBC subunits were co-expressed in yeast cells (Figure 1D). With this system, we could detect an interaction between the CBC and both Ars2 and PHAX. Ars2 contains a highly conserved C-terminal domain, which can be found in the fission yeast homolog of Ars2 (aa 636 to 871; Supplementary Figure 1D). We tested the N- and C-terminal domains separately in yeast two-hybrid assays, and the results indicated that the conserved C-terminal domain of Ars2 mediated its interaction with CBC.

The results presented above suggested that CBC, PHAX and Ars2 could form a complex. To prove this formally, we set out to reconstitute it from proteins purified *in vitro*. It was previously demonstrated that CBC and PHAX can associate and form a complex that we refer to as CBCP¹². Ars2 was shown to interact with the CBC in cell extracts¹⁷, but a direct physical interaction was not demonstrated. This is however an important point because dozens of proteins co-purify with the CBC even in RNase treated extracts¹⁹, and identifying the direct CBC partners is thus an essential step to better understand its mechanism of action. To this end, recombinant CBC was mixed with recombinant Ars2, and the mixture was added to m7-G resin. Ars2 and both CBC subunits co-purified on the resin, and when the eluate was analyzed by gel filtration, a major peak containing the CBC in complex with Ars2 was observed (Figure 2A). This demonstrates that CBC and Ars2 associate together to form a complex that we named CBCA. We then repeated a similar experiment using Ars2 and a preformed CBCP complex

(Figure 2B; ²⁰). All the four proteins co-purified on the m7G column. Upon gel-filtration, the eluted material migrated as a major peak that contained the four proteins in similar amounts, and which was well separated from the peaks produced by Ars2 alone or by the CBCP complex. Altogether, these results indicated that PHAX was associated with CBC and Ars2 in cells, and that a stable protein complex could be formed in vitro with recombinant proteins. This complex was thus referred to as CBCAP, for CBC, Ars2 and PHAX.

In vitro, Ars2 and PHAX bind independently to the CBC. In vivo however, PHAX appears to bind CBC together with Ars2. We thus tested whether Ars2 was required for the binding of PHAX to the CBC using the sensitive and quantitative LUMIER assay ²¹. PHAX was fused to Renilla luciferase, co-transfected in HeLa cells together with Flag-CBP20 in presence of siRNAs against Ars2 or control siRNAs. Anti-Flag immunopurification was performed, and luciferase activity was measured on the extracts and beads (Figure 2C). Depletion of Ars2 reduced the luciferase activity in the pellets relative to the input by more than 2 fold, indicating that in vivo, Ars2 stimulated the binding of PHAX to the CBC.

CBC and Ars2 associate with a similar set of capped RNAs, including pre-snRNAs, capped pre-snoRNAs and mRNAs

CBC and PHAX associate with m7G-capped snRNA and snoRNA precursors ^{10,12}, raising the possibility that Ars2 would do the same. To test this possibility, we performed a metabolic labeling in HeLa cells, in which we pulse-labeled RNAs with ³²P for optimal visualization of snRNA precursors. Extracts were then subjected to immunoprecipitation with anti-PHAX or control anti-HIV-1-p24 antibodies, and with a GST-tagged version of Ars2 or GST-MAP12 as control (Figure 3A). As expected, PHAX antibodies immuno-precipitated U3, U2, U1, U4 and U5 precursors, but not uncapped small RNAs like U6 and SRP. Although generally less efficient, GST-Ars2 immuno-precipitated a similar set of RNAs, including precursors of U2 and

U1. Mature U1, U2 and U5 were weakly detected in the GST-Ars2 pellets, but they were also present in the control IP and we were unable to demonstrate a specific association with these species (Figure 3A and unpublished data). To confirm the association of Ars2 with pre-snRNA, we created stable cell lines expressing tagged proteins. Single copy genes containing 3xFlag-tagged versions of Ars2 and CBP20 were introduced in the genome of HeLa cells by homologous recombination with the Flp-In system (Figure 3B). Western blots with anti-Flag antibodies revealed the presence of tagged proteins in the Flag cell lines (Supplementary Figure 2A). Furthermore, analysis with anti-Ars2 antibodies revealed that the amount of proteins was similar to that of the parental cells (Supplementary Figure 2A), indicating that the tagged proteins were expressed at near endogenous levels. To confirm the association of Ars2 with pre-snRNAs, we prepared extracts from the stable cell lines, immuno-precipitated the tagged proteins with anti-Flag beads, and analyzed the precipitated RNAs by RT-qPCR with amplicons specific for U2 precursors. The IP was efficient since the tagged protein was depleted from the extracts (Supplementary Figure 2B). In addition, when compared to the parental cells that have no tagged proteins, the Flag-Ars2 pellets were highly enriched in pre-U2 RNAs (40 fold; see below Figure 3E), confirming the results of the metabolic labeling experiment. Since it was previously demonstrated that CBC and PHAX bind capped pre-snRNA and pre-snoRNAs, these results demonstrated that the entire CBCAP complex associates with these RNA species.

PHAX is known to bind to small, capped non-coding RNAs^{10,12}, while Ars2 has been also involved in the biogenesis of miRNAs and histone mRNAs^{17,22}. It was thus likely that Ars2 would bind other RNAs besides pre-snRNAs and capped pre-snoRNAs. To define the entire set of RNAs associated with Ars2 and to determine whether these RNAs are bound in the context of CBCA, we analyzed the RNAs associated with Ars2 and CBC with microarrays and compared the two IPs. We first tested whether the capped RNAs were specifically enriched in the two IPs by comparing two equivalent groups of small non-coding RNAs, the m7G capped group that

contained mostly snRNAs and scaRNAs (68 species, median size of 201 nucleotides), and the uncapped group that contained intronic snoRNAs, pre-miRNAs and various pol III transcripts (408 transcripts, median size of 232 nucleotides). The enrichment in the IP was calculated by subtracting the \log_2 values of the specific IP to the \log_2 values of the control IP performed in the parental cells with no tagged protein. The "capped" group was enriched in the Ars2 and CBC IPs (median enrichment from 2.0 to 3.8 fold in linear scale, Supplementary Figure 2C), while the "uncapped" group was not enriched. These results indicated that both CBC and Ars2 bind specifically to capped RNAs.

Next, we analyzed the microarrays to find the entire set of RNAs bound by CBC and Ars2. Using a threshold equal to the mean plus 1.5 standard deviations (STD), we found that sets of 2003 and 1713 RNAs were enriched in the CBP20 and the Ars2 IP, respectively (Figure 3C). Interestingly, 61 % of the RNAs associated with Ars2 were also associated with CBC, indicating that the two proteins bound similar RNAs. In addition, taken as a whole, the RNAs enriched in the Ars2 set but not in the CBP20 set were still associated with CBP20, although below the threshold of 1.5 STD (mean \log_2 ratio over control IP of 0.43; pValue 10^{-16}). Likewise, the RNAs in the CBC set but not in the Ars2 one were also bound by Ars2 (\log_2 ratio of 0.39, pValue 10^{-16} ; Supplementary Figure 3A). To analyze this further, the two IP datasets were plotted against each other. A strong correlation could be seen (Pearson correlation coefficient of 0.8; Figure 3D), indicating that CBC and Ars2 associated in a very similar manner to RNAs, and thus that they most often associate in the context of the CBCA complex.

We then examined in detail the RNA families associated with these proteins. Visual examination revealed a number of small, capped non-coding RNAs such as snRNA, capped snoRNA and telomerase RNA, but also a high number of mRNAs. No particular GO term was enriched, many were however abundantly expressed mRNAs encoding housekeeping proteins such as histones and mitochondrial ribosomal proteins (see Supplementary Figure 3B,

Supplementary file). We next used RT-qPCR to verify some of the mRNA targets: CKS2, Cox7A2, RPL9 and RPS12. In agreement with the microarray data, all these mRNAs were enriched in the Flag-Ars2 IPs, albeit to various extent (Figure 3E). We then analyzed the Flag-Ars2 immuno-precipitates by RNase protection assays. A representative histone RNA, Hist1H4b, was enriched in the Ars2 IPs (Figure 3F). Since replication-dependent histone mRNAs follow a particular processing pathway without polyadenylation (reviewed in ²³) we also analyzed a canonical mRNA, RPS28. Using a probe covering the polyA tail, we found that polyadenylated RPS28 mRNAs were present in the Ars2 IPs (Figure 3F).

One surprising result of the microarray data was that we did not detect binding of CBC to all mRNAs as could have been expected. However, this may be due to technical limitations. The CBC binds mRNAs transiently, from the time the cap is added to the first round of translation ⁴. Thus, long-lived transcripts will spend most of their time without the CBC. Likewise, the CBC-bound fraction may be difficult to detect for low abundant mRNAs. It is thus possible that CBC binds all cellular capped RNAs as anticipated, but that we can detect only a subset of these interactions. Since Ars2 binds similar RNAs as the CBC, this would suggest that CBCA binds most or all capped RNAs in the cell nucleus. Alternatively, it is possible that Ars2 stabilizes the binding of CBC to a subset of capped RNAs, and that these are the events that we detect preferentially. In any case, our results demonstrate that CBC and Ars2 bind a similar large set of mRNAs, suggesting that these associations occur via the CBCA complex.

CBC and Ars2 depletions lead to transcriptional read-through of snRNA 3'-end processing sites

To define the function of Ars2 during RNA biogenesis, we knocked it down by RNA interference and analyzed the resulting changes in RNA abundance with tiling micro-arrays covering chromosomes 1 and 6, using a random-primed library. We first focused our attention on

snRNAs. There are four U1 genes on human chromosome 1 and visual inspection of the tiling arrays revealed an increase in signals downstream the region coding the mature RNA, as compared to a control siRNA (see below). Because its sequence is repeated over the genome, the U1 RNA itself was not present on the array. However, inspection of other snRNA loci such as U11 revealed that the increase in signal was specific for the region downstream of the mature snRNA (Supplementary Figure 4A). To provide a better quantification of this phenomenon, the U1 genes were aligned by their transcription start site and averaged (Figure 4A). The results confirmed an increased signal downstream the mature ends of U1 RNA. This suggested a defect in snRNA 3'-end processing, because the lack of recognition of 3'-processing signals is expected to lead to increased transcription through downstream sequences. Interestingly, we also detected an increased signal upstream of the U1 promoter. This is likely due to a promoter-associated RNA, or PROMPT^{24,25}, whose degradation is dependent on recruitment of the RNA exosome by CBC and Ars2 (see discussion and accompanying paper).

To confirm the defect in snRNA 3'-end processing upon Ars2 depletion, we developed a quantitative transcription read-through assay, in which the Firefly luciferase was inserted downstream of 3'-processing signals of snRNA genes (Figure 4B-D). The luciferase sequence possessed its own polyA signal, such that read-through transcripts could be processed and quantitated by luciferase activity. Variations in transfection efficiencies were corrected by co-transfecting a Renilla luciferase expression plasmid. This gene had the CMV promoter and SV40 polyA site, as the Firefly luciferase, thus factoring out effects at the level of mRNA 3'-end processing. This type of assay has been used previously and provides a sensitive and quantitative measure of read-through transcription²⁶. We constructed several reporters for snRNA 3'-processing based on U4. In these reporters, the U4 Sm site was mutated to prevent assembly with Sm proteins in the cytoplasm. Moreover, the RNA contained an additional short sequence to discriminate it from the endogenous U4 snRNAs. Because snRNA 3'-end processing requires the

snRNA promoter, the first plasmid contained the entire U4 gene including its own promoter²⁷. The other plasmids were negative controls that either transcribed the luciferase sequence directly from the U4 promoter, in absence of the U4 snRNA sequences and processing sites, or that transcribed the U4 snRNA and processing sites from the CMV promoter. These reporters were expressed in HeLa cells and various proteins were knocked-down to evaluate their effect on read-through transcription. We focused on CBCAP proteins and included some additional controls: CPSF3L, which is part of the integrator complex and is a positive control as it is the putative endonuclease that processes snRNAs 3'-ends²⁸; RRP40, which is part of the exosome and copurifies with Ars2 and CBC, and should indicate effects at the level of RNA stability²⁹; NELF-E, which directly binds the CBC and is part of NELF, a complex that is involved in snRNA 3'-end processing and also couples the cap with histone 3'-end processing^{15,30}.

The results showed a 2 to 5 fold increase in read-through signals upon depletion of Ars2, CBP20, CBP80 or CPSF3L (Figure 4B-D), while knock-downs of PHAX, RRP40 and NELF-E had no effect. Little effects were seen on the control reporters having the CMV or U4 promoters directly driving expression of Firefly luciferase (up to 1.5 increase in signal), indicating that transcription initiation contributed marginally to the effects observed. To confirm the data obtained with U4, we constructed a similar reporter based on U3. Again, we found that CBC, Ars2 and CPSF3L depletions induced read-through transcription past the U3 3'-processing signals (Supplementary Figure 4B). Altogether, these effects indicated a role for CBC and Ars2 in the 3'-end formation of snRNAs and capped pre-snoRNAs.

To confirm these data, we performed RT-qPCR analyses using amplicons located either upstream or downstream the 3'-end of U4 snRNA, for both the luciferase reporter and the endogenous U4 RNA (Figure 4E-F). Although the effects varied slightly for each protein, we consistently observed more read-through transcripts upon knock-downs of CBC, Ars2 and CPSF3L (1.3 to 3 fold). These data were further supported by Pol-II chromatin immuno-

precipitation (ChIP) experiments on two endogenous U4 genes, in which 1.2 to 2 fold more Pol-II could be detected downstream the U4 3'-end upon depletion of Ars2 (Figure 4G). Altogether, these data demonstrate that CBC and Ars2 stimulate snRNA 3'-end processing and that their depletion induce transcriptional read-through.

CBC and Ars2 knock-downs lead to read-through of histone 3'-end processing signals

Since Ars2 binds many mRNAs in addition to pre-snRNAs, we analyzed the tiling array data for defects in mRNA biogenesis upon Ars2 depletion. Visual inspection revealed an increase in signal in regions downstream histone mRNAs (Figure 5A). Histone genes were aligned with respect to their TSS and signals from all the 75 histone genes were averaged together. The data showed a decrease in signal in the coding region of histone genes, while the signal downstream 3'-processing sites increased. Most histones mRNAs are not polyadenylated but follow an alternate U7-dependent 3'-processing pathway where cleavage is not followed by polyadenylation ²³. Because Ars2 depletion did not affect U7 snRNA levels (data not shown), and because it bound directly to histone mRNAs, these data suggested a direct role of Ars2 in histone 3'-end processing. To analyze this further, we selected the processing site of the histone Hist1H3b gene and inserted it between the CMV promoter and the luciferase coding region to create a read-through reporter similar to the one previously used for U4 snRNA. As a positive control, we knocked-down FLASH, a protein playing an essential role during histone 3'-end formation ³¹. Remarkably, luciferase activity increased 1.8 to 2.8 fold following depletion of CBP20, CBP80, Ars2 or FLASH, while the knock-downs of NELF-E, CPSF3L, RRP40 and PHAX had no effects (Figure 5B). This confirmed that Ars2 was involved in histone 3'-end formation, as previously proposed ²², and it further indicated that this occurred in the context of the CBCA complex (see discussion below).

CBC and Ars2 knock-downs lead to read-through of canonical and cryptic polyA sites

The previous experiments indicated a role for CBC and Ars2 in the 3'-end processing of two types of genes. Since the CBC and Ars2 bind many mRNAs, these data prompted us to test whether these factors would also be involved in the processing of pA sites. To test this possibility, we inserted various pA sites between the CMV promoter and the Firefly luciferase sequence to perform read-through assays. We chose two polyA sequences that were derived from genes strongly bound by CBC and Ars2: CENPW and CNIH4. When read-through transcription was measured by the luciferase assay, a 3 to 6 fold increase was seen for the CENPW and CNIH4 reporters after knockdown of CBC or Ars2 (Figure 5C and S5A). No effect was observed for PHAX, CPSF3L, FLASH, RRP40 and NELF-E, even though CPSF3L and FLASH generated a positive signal with the U4 and Hist1H3b reporters, respectively. To assess the specificity of the effects, we mutated the AAUAAA sequences present in the CENPW polyA region. The mutations effectively induced transcriptional read-through as they increased Firefly luciferase activity by about 6 fold. Little or no effect of CBC and Ars2 knock-downs was seen with this mutant construct (Figure 5C), indicating that the effects of CBC and Ars2 required the AAUAAA signals.

In addition to functional polyA sites that promote 3'-processing at the end of mRNAs, the genome contains many cryptic sites that can lead to premature cleavage and polyadenylation (PCAP). These often occur in the first intron of coding mRNAs and upstream of transcription start sites in antisense orientation³²⁻³⁴. This limits antisense transcription whereas the cryptic sites in the sense orientation are suppressed by splice donor sequences. The NR3C1 gene is a model for this phenomenon³³. It contains a cryptic pA site in its first intron, which is normally repressed by the binding of U1 snRNP to the splice donor site, but that can be activated by U1 knock-down or by mutating the splice donor sequence. To test whether CBC and Ars2 would promote processing at this cryptic polyA sequence, we cloned it in the luciferase system with

part of the first NR3C1 exon. Upon knock-down of CBP20, CBP80 or Ars2, no effect was detected with the wild-type sequence in which the cryptic pA site is suppressed (Figure 5D). As expected, mutation of the splice donor activated the cryptic pA site as this decreased luciferase activity by 7-fold. With this mutant, knock-downs of CBC and Ars2 induced 2-3 fold increase in luciferase activity, and this effect was linked to cleavage and polyadenylation because it was abolished by further mutating the four AAUAAA-like sequences present within the NR3C1 intron (Figure 5D and S5B). Next, we confirmed that the increases in luciferase signals were due to read-through transcription. To this end, we analyzed RNAs by RT-qPCR with amplicons located upstream and downstream the polyA sites, for both the CENPW and NR3C1 reporter plasmids (Figure 5E and S5C). This analysis was also performed for the endogenous CENPW gene to exclude the possibility that the effects of CBC and Ars2 were due to the plasmid system used (Figure 5E). In all cases, a 1.5-2 fold increase in read-through transcripts was measured upon Ars2 or CBP20 depletion, confirming that the absence of these proteins induced transcriptional read-through, and thus that they were stimulating 3'-end processing at polyadenylation sites.

Ars2 links the CBC to the cleavage factor CLP1

The CBC has long been known to stimulate cleavage at polyA sites, but its mechanism of action has remained obscure. In addition, CBC also enhances 3'-end processing of histone mRNAs¹⁵, and results shown here indicate that it also increase 3'-end processing of snRNAs. To understand how Ars2 could mediate these effects of the CBC on 3'-end processing, we characterized its partners by quantitative SILAC proteomics. We used a GFP-tagged construct stably expressed in U2OS cells, and the experiment was performed in absence of RNase treatment in order to also capture weakly interacting proteins. Analysis of the pull-downs indicated that Ars2 co-purified with the CBC and PHAX as expected (Figure 6A). About 20 other proteins were also enriched,

among which a number of chaperones and α -importins. Only three proteins were directly related to RNA processing: CLP1, hnRNPUL1, and EWSR1. CLP1 was especially appealing because it has been reported to play a key function during cleavage at polyA sites. First, CLP1 co-purifies with the TSEN complex, which is involved in both tRNA splicing and cleavage at mRNA polyadenylation sites³⁵. Second, CLP1 is present in purified fractions of the cleavage factor CFII(m), where it interacts with the Pol-II CTD binding protein PCF11 and plays an essential role by bridging CPSF with CFI(m)³⁶. To strengthen the link between CBC, Ars2 and CLP1, we performed a series of co-IP experiments. First, we over-expressed a GFP-tagged version of CLP1 in HeLa cells expressing Flag-tagged versions of Ars2 and CBP20 and immunoprecipitated CLP1 with anti-GFP antibodies. In both untreated and RNase-treated extracts, we detected an association of CLP1 with Flag-Ars2 and Flag-CBP20 (Figure 6B). Likewise, we detected an RNA-independent interaction between CLP1 and Pol-II (Supplementary Figure 6). Next, we repeated the experiment using a stable HeLa cell line expressing GFP-CLP1 from a BAC containing the entire CLP1 gene including all its natural regulatory sequences (Figure 6C). In these conditions, GFP-CLP1 is expressed at native levels and its interactions thus more likely reflect the situation of the endogenous protein³⁷. We could detect a robust signal for both the endogenous Ars2 and CBP20 in the GFP-CLP1 pellets. However, this interaction was disrupted by RNase treatment of the extracts. Altogether, these data suggest that the association between Ars2 and CLP1 is mediated by protein-protein interaction but is stimulated or stabilized by RNA under the low expression levels of the endogenous proteins.

To analyze the functional relevance of this interaction, we tested the effects of depleting CLP1 and PCF11 on read-through transcription using the various luciferase reporters. The depletion of CLP1 was modest (Supplementary Figure 7). Nevertheless, knock-downs of either CLP1 or PCF11 induced a two to three-fold increase in luciferase activity as compared to the control plasmids, for the U4, Hist1H3b and CENPW reporters (Figure 6D). CLP1 and PCF11

were not previously known to be involved in 3'-end processing of snRNAs and histone mRNAs, and the similarity of the effects between CBC and Ars2 on one side, and CLP1 and PCF11 on the other, suggests that the interaction between Ars2 and CLP1 is functionally relevant and mediates at least part of the effect of the cap on 3'-end processing.

The effects of CBC and Ars2 depend on the distance between the cap and the processing site

The physical interaction between the CBC and CLP1 prompted us to test whether the distance between the cap and the tested 3'-end processing signals plays a role for the stimulatory effect of the CBC and Ars2. To test this idea, we created new luciferase reporters in which the distance between the cap and the processing sites was increased (Figure 7A). For the U4 reporters, the effects of CBC, Ars2 and CPSF3L knockdowns decreased when the distance moved from 225 to 575 nt, and was abolished when it was 725 nt. Thus, we observed a decrease in both the strength of the processing signal and the ability of CBC and Ars2 to stimulate it. Similarly, the effect of CBC, Ars2 and FLASH siRNAs decreased when the distance between the Hist1H3b histone processing site and the cap increased from 260 to 1260 nt. In the case of the NR3C1 reporter in which the cryptic pA site was active, the effect of CBP20 and Ars2 siRNAs were decreased by more than 2 fold when a 1kb spacer was inserted between the cap and the polyadenylation site (Figure 7A). Surprisingly, the effect of depleting of CBP80 was not significantly affected by the spacer. Since CBP20 is the subunit that contacts directly the cap, this data suggested that the cap, together with Ars2, stimulated preferentially cap-proximal polyadenylation sites, while CBP80 could also stimulate cap-distal sites, perhaps by a mechanism independent of Ars2 and that can be uncoupled from binding to the cap. Altogether, these results indicate that the effects of the cap and Ars2 are dependent on gene length and that they preferentially stimulate cap-proximal processing sites.

DISCUSSION

The CBC has long been known to stimulate mRNA splicing and 3'-end formation. However, its partners in these processes and its mechanisms of action have remained elusive. Here, we show that recombinant Ars2 and CBC can form a stable complex in vitro (CBCA), that the CBC and Ars2 bind very similar sets of capped RNAs in vivo and that they stimulate a variety of 3'-end processing signals. This indicates that Ars2 is a major effector of the CBC in human cells.

The CBC and Ars2 couple snRNA 3'-end processing with export

The CBC stimulates many steps of RNA maturation, and it is thus ideally placed to integrate and coordinate them, but whether this occurs is not known. In this study, we have identified the CBCAP complex, composed of CBC, Ars2 and PHAX. In vitro reconstitution experiments show that CBC can form independent complexes with either PHAX (CBCP) or Ars2 (CBCA), but in vivo data indicate that most of PHAX binds CBC together with Ars2. Our results show that Ars2 stimulates both snRNA 3'-end processing and the association of PHAX with the CBC (see model in Figure 7B). It thus physically and functionally couples snRNA maturation with export. Ars2 may directly stabilize PHAX on the CBC. Alternatively, PHAX binding may be stimulated by snRNA 3'-end formation, and Ars2 may thus act indirectly. Export of mRNA was previously linked to 3'-end processing in yeast³⁸, but this was not known for non-coding RNAs that use a distinct export pathway.

CBCA stimulates 3'-end processing of many RNA families

We observed that depletion of CBC or Ars2 leads to transcriptional read-through of pA sites, indicating that CBCA can stimulate mRNA 3'-end processing. Interestingly, we demonstrated that CBC and Ars2 also stimulated 3'-end processing of snRNAs and replication-dependent histone genes. These families follow distinct maturation pathways, indicating that Ars2 is a

general and versatile adaptor that links the cap to a number of RNA 3'-end processing machineries. This result corroborates a proteomic analysis of the proteins that assemble on a short RNA carrying a model AAUAAA cleavage and polyadenylation site ³⁹. This study detected a very large complex containing not only the canonical components of the cleavage machinery, but also the integrator complex that processes snRNAs. Thus, the 3'-end processing machinery may exist as a "super complex" able to process different types of RNA. This machinery may be correctly positioned and stabilized by cis-acting RNA sequences together with the CBC and Ars2.

Ars2 has been previously implicated in histone 3'-end formation, but it was unclear whether this effect was related to the CBC as its mechanism of action was not characterized ²². Our data indicate that the effect of CBC and Ars2 are linked and that they act through the cleavage factor CLP1. First, CBC and Ars2 co-purify with CLP1, and second, the depletion of these three proteins similarly affects 3'-end processing of snRNAs, histone mRNAs, and polyadenylated sites. Our results thus suggest a model in which the CBC recruits CLP1 through Ars2 (see Figure 7B). In this regard, it is interesting that the proteomic study cited above used an uncapped RNA to purify cleavage factors and failed to detect CLP1 ³⁹. Possibly, recruitment and stable binding of this factor requires binding of CBC and Ars2 to capped RNAs. This mechanism would also be consistent with current data on CBC. Indeed, although the mechanism of action of the CBC on polyadenylation sites remained unclear for many years ⁹, it was shown not to involve CPSF or CstF, which are major cleavage and polyadenylation complexes that do not contain CLP1.

CLP1 is an essential component of the cleavage factor CFII(m), where it interacts with the CTD-binding protein PCF11 ³⁶. Interestingly, it is also part of the TSEN complex, which is required for both tRNA splicing and cleavage at mRNA polyadenylation sites ³⁵. It is currently unclear whether TSEN associates with CFII(m) and whether these activities of CLP1 in

mediating mRNA 3'-end cleavage are related or not. Depletion of PCF11 affected 3'-end processing of snRNAs, histone mRNAs, and polyadenylated mRNAs, consistent with a role for this factor in the effects of CLP1. We attempted to deplete TSEN2 but this was inefficient and the role of this factor could thus not be properly addressed (data not shown). Taken together, our data indicate a broad role of CLP1 and PCF11 in 3'-processing at a variety of signals, and suggest that CBC and Ars2 help to recruit these factors.

The CBC stimulates histone mRNA 3'-end formation through multiple partners

The CBC was previously shown to be involved in histone 3'-end formation, and it was demonstrated to function as part of a ternary complex containing NELF-E and SLBP¹⁵. NELF induces polymerase pausing at the promoter and is part of the so-called elongation check-point. Thus, NELF was proposed to recruit the CBC on promoter-paused polymerases to improve downstream RNA processing once the polymerase is released. We have been unable to detect an effect of NELF-E depletion in our luciferase read-through assays, perhaps because the CMV and histone promoters differ in their ability to recruit NELF or to assemble the NELF-CBC-SLBP complex. This however indicates that the effects of CBC and Ars2 that we observe are independent from the interaction of the CBC with NELF. It is interesting to note that Ars2 also interacts physically with FLASH⁴⁰, a protein that plays an essential role in histone 3'-end formation. These data thus suggest that the CBC may stimulate histone mRNA processing by several ways: NELF/SLBP, Ars2/CLP1, or Ars2/FLASH. These factors have different targets and may thus act synergistically.

SnRNA 3'-processing also requires NELF³⁰. As proposed for histone genes, the role of NELF may be to ensure CBC loading before polymerases are released from the promoter. In the case of short genes such as histones and snRNAs, elongation may take as little as 2 seconds and it would thus be especially important to ensure that the 3'-processing machinery is loaded when

polymerases are released from the promoter. Such a rapid kinetic may also explain the requirement on a cognate promoter for snRNA 3'-end formation²⁷.

CBC, Ars2 and discrimination of cryptic transcripts

Recent progresses in sequencing techniques have indicated that a major fraction of the genome is transcribed. One fundamental question that arose from these studies is to understand the specific features of the transcriptional events that lead to the production of an mRNA. Premature cleavage and polyadenylation sites and U1 snRNP appear to be important factors that discriminate functional from cryptic transcripts. Indeed, polyadenylation sites are present close to transcriptional start sites and trigger premature transcript termination in the antisense orientation, while they are suppressed by splice donor sequences in the sense orientation³²⁻³⁴. Interestingly, Ars2 knock-downs lead to PROMPT stabilization (Figure 4A, see accompanying paper). In addition, CBC and Ars2 appear to specifically stimulate premature pA signals, since their effect is sensitive to the distance from the cap and is suppressed by splice donor sequences (this study). This suggests that CBC and Ars2 may play an important role in selecting functional from cryptic transcripts, and it further indicates that the distance of the polyA site from the cap is an important feature that distinguishes mRNAs. Co-transcriptional measurement of gene length would be analogous to the situation in yeast, where the NRD1 complex can trigger transcription termination only at a short distance from the transcription start site⁴¹, thanks to its interaction with S5-phosphorylated Pol-II CTD^{41,42}. Here, Ars2 and CBC are recruited by the cap, and increasing the length of the transcript may for instance simply decrease their probability to interact with the elongating polymerase (see model in Figure 7C).

CBC and Ars2 have been recently shown to co-purify with NEXT, a nucleoplasmic co-factor of the exosome whose interaction is functionally relevant as knock-down of CBC and Ars2 leads to the stabilization of RNAs normally degraded by the exosome⁽²⁹⁾; see

accompanying paper). CBC and Ars2 thus appear to form a key node in RNA processing, which physically connects machineries involved in 3'-end formation, degradation and export. This is remarkable because it places the CBC and Ars2 at the intersection of productive and degradative pathways, raising the fascinating possibility that they select the fate of newly synthesized RNAs. For instance, the CBC could ensure that a nascent RNA would be exported when cleaved at the proper processing site, and degraded when the polymerase would read through this site and process the RNA at a downstream signal. In agreement with this scenario, Ars2 stimulates the binding of PHAX to 3'-end processed pre-snRNAs, while NEXT associates with snRNA and histone mRNA readthrough transcripts and is involved in their degradation (see accompanying paper). In the future, a key question will thus be to understand the specific features that lead the CBC to route RNAs for productive or degradative pathways.

Acknowledgements.

We thank the Functional Proteomics Platform (FPP, Montpellier Languedoc-Roussillon, France) for the use of their instruments. We thank Ina Poser for the gift of CLP1-GFP BAC cell line, Hervé Le Hir, and Maiken S. Kristiansen for critical reading of the manuscript. D. L. had a fellowship from the FRM, M. H. from the MNRT, F. P. from La Ligue Contre le Cancer. This work was funded by grants from the Danish National Research Foundation (grant D NRF58) to T.H.J., and from the CNRS, l'ARC and La Ligue Contre Le Cancer to E. B. The authors declare they have no conflicts of interest.

Methods

Plasmids and cell culture

Hela cells were cultivated in DMEM with antibiotics and 10% fetal calf serum (FCS). Stable Flp-In isogenic Hela cells expressing the Flag-tagged proteins were obtained from pcDNA5 vectors co-transfected with a Flippase expression vector and hygromycin selection of expressing clones as recommended by the manufacturer (Invitrogen). Hela cells containing CLP1-GFP-tagged BAC were described previously ³⁷. DNA cloning was performed by standard techniques and with the Gateway™ system (Invitrogen). Detailed maps and sequences are available upon requests.

Immuno-precipitation, LUMIER assays and siRNA treatments

For IPs, cells were extracted in HNTG buffer (20 mM HEPES, pH 7.9, 150 mM NaCl, 1% Triton, 10% glycerol, 1 mM MgCl₂, 1 mM EGTA, and protease inhibitors) for 30 min at 4°C. Cellular debris were removed by centrifugation (10 min at 9000 g). Extracts were put on antibody-coated beads for 2h at 4°C, washed twice in HNTG and three times in PBS, resuspended in 1% SDS at RT, and finally in Laemli buffer. When indicated, RNase A was added at 60 µg/ml. For SILAC, we used cryogrinded extracts (see Supplementary Methods). For LUMIER assays, PHAX was fused to Renilla luciferase and co-transfected with Flag-CBP20. Luciferase activity was measured in extracts and pellets with the dual luciferase kit (Promega). SiRNA and plasmids were transfected for 48h with JetPrime (Polypus), or with Lipofectamine and Plus reagent (Invitrogen), as recommended by the manufacturers.

Metabolic labelling and RNA analyses

HeLa cells were washed in phosphate-free DMEM containing 10% dialyzed FCS, incubated for 1h in this buffer, and then 2.5 h in the same buffer plus 0.5 mCi ³²P. Cells were lysed in

HNTG, immuno-purified with the indicated antibodies and RNAs were purified in Trizol (InVitrogen). RT-qPCR and micro-arrays (Affymetrix) were performed as described previously ⁴³.

In vitro reconstitution of CBCAP complex

PHAX (1-318) and CBC Δ CC Δ NLS were produced as previously described ²⁰. His-tagged ARS2 (residues147-871) was cloned in ProExHTb expression vectors, expressed in E. Coli BL21Star(DE3) grown at 20°C and purified on Ni²⁺ resin after induction. Complexes were formed by mixing purified proteins and then re-purified on m7-G resin. Following elution with m7-G, proteins were loaded on a Superdex S200 column. Fractions of interest were analyzed by SDS-PAGE and gels were stained with coomassie blue.

References

1. Kelly, S. & Corbett, A. Messenger RNA export from the nucleus: a series of molecular wardrobe changes. *Traffic* **10**, 1199-208 (2009).
2. Luna, R., Gaillard, H., González-Aguilera, C. & Aguilera, A. Biogenesis of mRNPs: integrating different processes in the eukaryotic nucleus. *Chromosoma* **117**, 319-31 (2008).
3. Perales, R. & Bentley, D. "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol Cell* **36**, 178-191 (2009).
4. Topisirovic, I., Svitkin, Y., Sonenberg, N. & Shatkin, A. Cap and cap-binding proteins in the control of gene expression. *Wiley Interdiscip Rev RNA*. **2**, 277-98 (2011).
5. Rasmussen, E. & Lis, J. In vivo transcriptional pausing and cap formation on three *Drosophila* heat shock genes. *Proc Natl Acad Sci U S A*. **90**, 7923-7 (1993).
6. Cho, E., Takagi, T., Moore, C. & Buratowski, S. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **11**, 3319-3326 (1997).
7. McCracken, S. et al. 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev.* **11**, 3306-3318 (1997).
8. Izaurralde, E. et al. A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78**, 657-668 (1994).
9. Flaherty, S., Fortes, P., Izaurralde, E., Mattaj, I.W. & Gilmartin, G. Participation of the nuclear cap binding complex in pre-mRNA 3' processing. *Proc Natl Acad Sci U S A*. **94**, 11893-8 (1997).
10. Boulon, S. et al. PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. *Mol. Cell* **16**, 777-787 (2004).

11. Cheng, H. et al. Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* **127**, 1389-400 (2006).
12. Ohno, M., Segref, A., Bachi, A., Wilm, M. & Mattaj, I.W. PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell* 187-198 (2000).
13. Kim, K. et al. A new MIF4G domain-containing protein, CTIF, directs nuclear cap-binding protein CBP80/20-dependent translation. *Genes Dev.* **23**, 2033-2045 (2009).
14. Lewis, J., Izaurralde, E., Jarmolowski, A., McGuigan, C. & Mattaj, I. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev.* **10**, 1683-98 (1996).
15. Narita, T. et al. NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs. *Mol. Cell* **26**, 349-365 (2007).
16. Ohno, M., Segref, A., Kuersten, S. & Mattaj, I.W. Identity Elements Used in Export of mRNAs. *Molecular Cell* **9**, 659-671 (2002).
17. Gruber, J.J. et al. Ars2 Links the Nuclear Cap-Binding Complex to RNA Interference and Cell Proliferation. *Cell* **138**, 328-339 (2009).
18. Boulon, S. et al. HSP90 and its R2TP/Prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II. *Mol Cell.* **39**, 912-24 (2010).
19. Pabis, M. et al. The nuclear cap-binding complex interacts with the U4/U6{middle dot}U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. *RNA In press*(2013).
20. Mourão, A., Varrot, A., Mackereth, C.D., Cusack, S. & Sattler, M. Structure and RNA recognition by the snRNA and snoRNA transport factor PHAX. *RNA* **16**, 1205-1216 (2010).
21. Barrios-Rodiles, M. et al. High-throughput mapping of a dynamic signaling network

- in mammalian cells. . *Science* **307**, 1621-1625 (2005).
22. Gruber, J. et al. Ars2 promotes proper replication-dependent histone mRNA 3' end formation. *Mol Cell*. **45**, 87-98 (2012).
 23. Dominski, Z. & Marzluff, W. Formation of the 3' end of histone mRNA: getting closer to the end. . *Genes Dev*. **396**, 373-390 (2007).
 24. Preker, P. et al. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851-1854 (2008).
 25. Seila, A. et al. Divergent transcription from active promoters. *Science* **322**, 1849-51 (2008).
 26. Chen, J. et al. An RNAi screen identifies additional members of the Drosophila Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation. *RNA* **18**, 2148-2156 (2012).
 27. Hernandez, N. & Weiner, A. Formation of the 3' end of U1 snRNA requires compatible snRNA promoter elements. *Cell*. **47**, 249-58 (1986).
 28. Baillat, D. et al. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265-76 (2005).
 29. Lubas, M. et al. Interaction profiling identifies the human nuclear exosome targeting complex. *Mol Cell*. **43**, 624-37 (2011).
 30. Egloff, S., O'Reilly, D. & Murphy, S. Expression of human snRNA genes from beginning to end. *Biochem Soc Trans*. **36**, 590-4 (2008).
 31. Yang, X., Burch, B., Yan, Y., Marzluff, W. & Dominski, Z. FLASH, a proapoptotic protein involved in activation of caspase-8, is essential for 3' end processing of histone pre-mRNAs. *Mol Cell*. **36**, 267-78 (2009).
 32. Almada, A., Wu, X., Kriz, A., Burge, C. & PA., S. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **In press**(2013).

33. Kaida, D. et al. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664-8 (2010).
34. Ntini, E. et al. Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *NSMB In Press*(2013).
35. Paushkin, S., Patel, M., Furia, B., Peltz, S. & Trotta, C. Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3' end formation. . *Cell* **117**, 311-321 (2004).
36. de Vries, H. et al. Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J.* **19**, 5895-5904 (2000).
37. Poser, I. et al. BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Methods* **5**, 409-415 (2008).
38. Hilleren, P., McCarthy, T., Rosbash, M., Parker, R. & Jensen, T. Quality control of mRNA 3'-end processing is linked to the nuclear exosome. *Nature.* **413**, 538-42 (2001).
39. Shi, Y. et al. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell.* **33**, 365-76 (2009).
40. Kiriya, M., Kobayashi, Y., Saito, M., Ishikawa, F. & Yonehara, S. Interaction of FLASH with arsenite resistance protein 2 is involved in cell cycle progression at S phase. *Mol Cell Biol.* **29**, 4729-41 (2009).
41. Gudipati, R., Villa, T., Boulay, J. & Libri, D. Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nat Struct Mol Biol.* **15**, 786-94 (2008).
42. Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S. & Meinhart, A. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol.* **15**, 795-804 (2008).

43. Pradet-Balade, B. et al. CRM1 controls the composition of nucleoplasmic pre-snoRNA complexes to licence them for nucleolar transport. *EMBO J.* **30**, 2205-2218 (2011).

Figure Legends.

Figure 1: The CBCAP complex contains PHAX, CBP20, CBP80 and Ars2.

(A) Proteins co-precipitated with PHAX from HeLa nuclear extracts were eluted in mild conditions, separated by 5-20% SDS-PAGE, stained with coomassie blue and analyzed by mass spectrometry. IP: anti-PHAX immunoprecipitation; Ct: control IP with anti-p24 antibody. Arrows: protein identified by mass-spectrometry. PHAX co-migrated with the non-specific band indicated with a star. (B) Co-purification of Ars2 with PHAX from nuclear and cytoplasmic extracts of HeLa cells. Extracts were treated with RNAse when indicated. Input: 5% of pellets. Star: non-specific band revealed by the antibody. (C) Co-recruitment assay to analyze interaction between PHAX and Ars2 in living cells. U2OS cells bearing a large LacO repeat were transfected with the indicated plasmids and observed by fluorescent microscopy. The Laci fusions also contained mCherry. (D) Yeast two-hybrid analysis. The indicated strains were mated and selected on -A-L-T for diploids, and on -A-L-T-H to test for interactions between the indicated proteins. Ars2-Ct: aa 502-871.

Figure 2: The CBCAP complex can be reconstituted in vitro and Ars2 contributes to the association of PHAX with CBC in vivo.

(A) Recombinant CBC complex was mixed with recombinant Ars2 and purified on m7G resin (left part of the gel). FT: flow-through; elution: material eluted with m7G. Eluted material was loaded on a gel filtration column (top graph). Proteins from each fraction were analyzed by SDS-PAGE and stained with Coomassie blue. Stars: Ars2 degradation products. (B) Same experiment as in A, except that a CBC-PHAX complex (bottom left panel) was incubated with recombinant Ars2. The gel filtration profiles are shown for each complex. Fractions from the mixture containing CBC-PHAX and Ars2 were analyzed by SDS-PAGE and stained with Coomassie blue (right gel). (C) LUMIER assays to evaluate the role of Ars2 in the binding of

PHAX to the CBC in vivo. Left: schematic of the assay. Right: LUMIER IPs. HeLa cells were transfected with Flag-CBP20 and PHAX-RLuc in presence of the indicated siRNA, and luciferase activity was measured in extracts and in the anti-Flag pellets. IP/Input ratios were normalized to that of the control siRNA.

Figure 3: Identification of the RNAs bound by the CBC and Ars2.

(A) Metabolic labelling of HeLa cells. The lane labelled "PHAX IP" correspond to an extract that was immunoprecipitated with anti-PHAX monoclonal antibodies (lane "PHAX mABs") or an anti-p24 as control (lane "Ct mABs"). Lanes labelled "GST-MAp12" and "GST-Ars2" were from cells transfected with GST-Ars2 and GST-MAp12 as control, and immunopurified with glutathione beads. Input: 5% of pellets. (B) Construction of isogenic Flp-In cell lines expressing 3xFlag-tagged CBP20 and Ars2, and schematic of microarray experiment. (C) Comparison of RNAs associated with 3xFlag-CBP20 (2003 RNAs with enrichment greater than the mean plus 1,5 STD) and 3x-Flag-Ars2 (1713 RNAs with enrichment greater than the mean plus 1,5 STD). (D) Flag-CBP20 and Flag-Ars2 IPs were plotted against each others and the Pearson correlation coefficient is indicated in parenthesis. (E) RT-qPCR analysis of RNAs purified from 3x-Flag-Ars2 IPs. IPs were normalised over GAPDH mRNA, which is not enriched in any of the IP in the microarray experiment. Values are fold enrichment over a control IP performed in the parental cell line that expresses no tagged proteins. (F) RNase protection analysis of RNAs associated with 3x-Flag-Ars2, using a probe hybridizing to the 3' end of Hist1H4b mRNA (covering 90 nucleotides, left), or to polyadenylated RPS28 mRNAs (covering 80 to 110 nucleotides depending on the length of the polyA tail, right). Mock lanes: probes incubated without cellular RNAs. Ct: control IP from the same extract and with beads lacking the anti-Flag antibody. Inputs: 5% of pellets.

Figure 4: CBC and Ars2 are involved in 3'-end processing and transcription termination of snRNA genes.

(A) HeLa cells were depleted in Ars2 and the resulting RNAs were analyzed by tiling microarrays covering chromosome 1 and 6. Four U1 genes were aligned by their transcription start site and averaged. Transcription goes from left to right. (B-D) Luciferase assays for read-through transcripts. HeLa cells were transfected with the depicted plasmids and siRNAs, and Firefly luciferase activity was measured and expressed as ratio over a co-transfected Renilla reporter plasmid. B: control plasmid with the CMV promoter driving expression of U4 snRNA and processing signals; C: test plasmid with the entire U4 gene upstream of Firefly luciferase; D: control plasmid with the U4 promoter immediately upstream of Firefly luciferase. (E-F) RT-qPCR analyses. HeLa cells were transfected with the indicated plasmid and siRNAs. Read-through transcripts that passed U4 processing sites were measured by RT-qPCR and expressed as ratio of an upstream amplicon. Both the U4 luciferase reported plasmid (E), and endogenous U4 were analyzed (F). (G) Pol II Chip was performed on HeLa cells treated with the indicated siRNA. Fold enrichment were normalized to the values obtained with the eGFP control siRNA.

Figure 5: CBC and Ars2 knock-downs lead to read-through of 3'-processing site of non-polyadenylated histone mRNAs and polyadenylated mRNAs.

(A) HeLa cells were depleted in Ars2 and the RNAs were analyzed by tiling microarrays. The 75 histone genes were aligned by their transcription start site and averaged. Transcription goes from left to right. (B) Luciferase assays for readthrough of Hist1H3b processing site. HeLa cells were transfected with the depicted plasmids and siRNAs, and Firefly luciferase activity was measured and expressed as ratio over a co-transfected Renilla reporter plasmid. (C) Luciferase assay for readthrough of CENPW cleavage and polyadenylation site. Left:

Hela cells were transfected with the CENPW reporter plasmid together with the indicated siRNA. Right panel: mutations of the AAUAAA hexamers of CENPW induced read-through transcription. (D) Same as in C, except that the cryptic polyA sequence of the first intron of the NR3C1 gene were used. Right panel: mutation of the splice donor sequence reduced read-through transcription, while further mutating the AAUAAA hexamers increased it. iPAS: intronic polyadenylation site; SD: splice donor site. (E) RT-qPCR analyses. Hela cells were transfected with the indicated plasmid and siRNAs. Read-through transcripts that passed CENPW processing sites were measured by RT-qPCR and expressed as ratio of an upstream amplicon. The luciferase reporter and endogenous CENPW gene were analyzed.

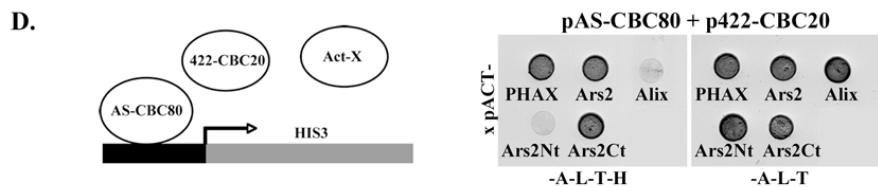
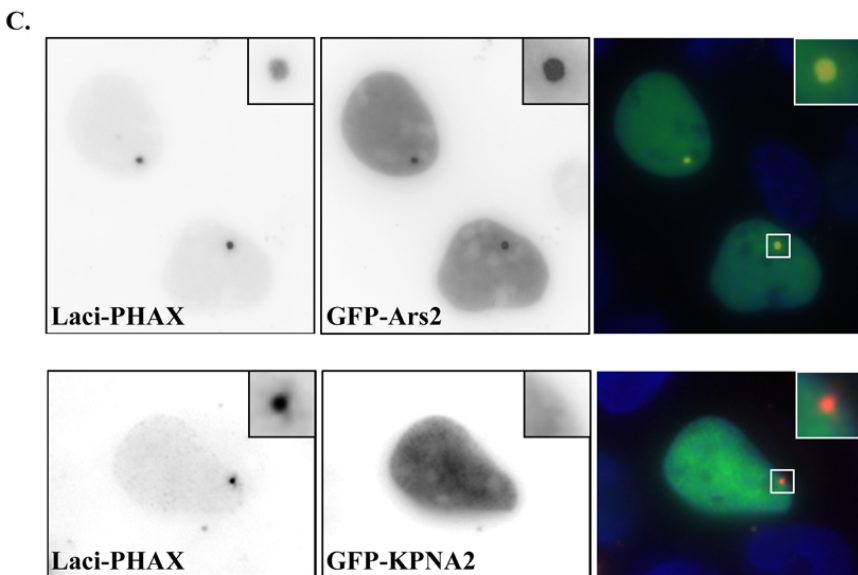
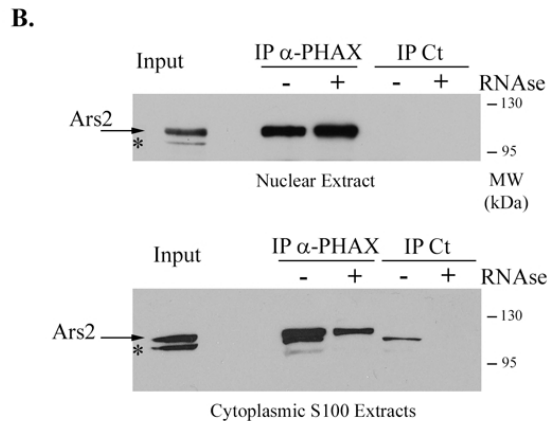
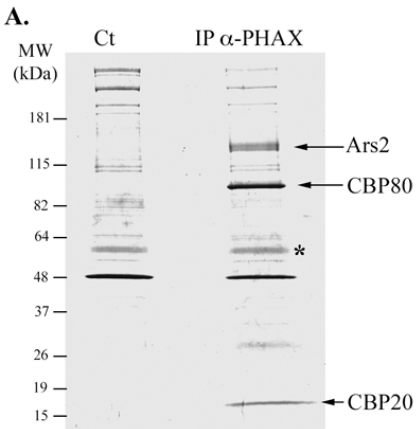
Figure 6: CBC and Ars2 associate with the cleavage factor CLP1.

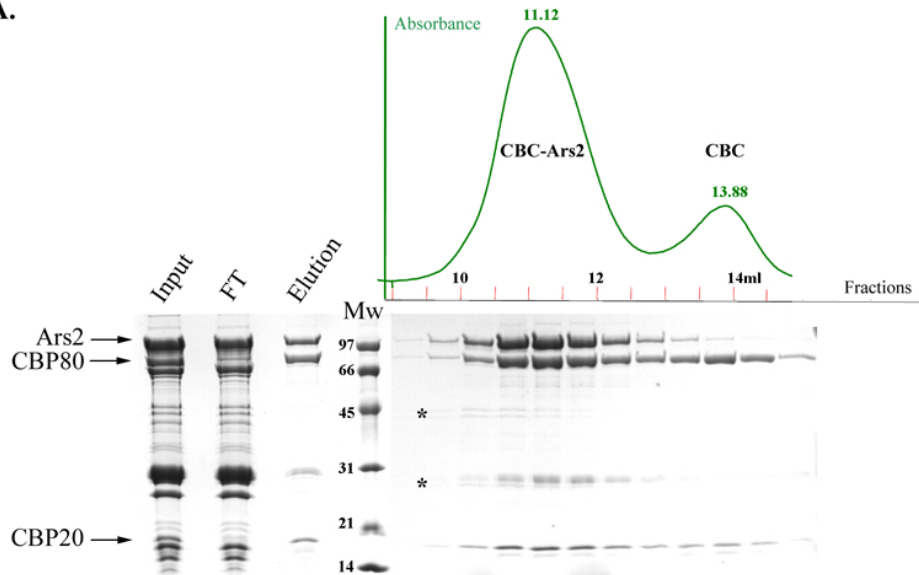
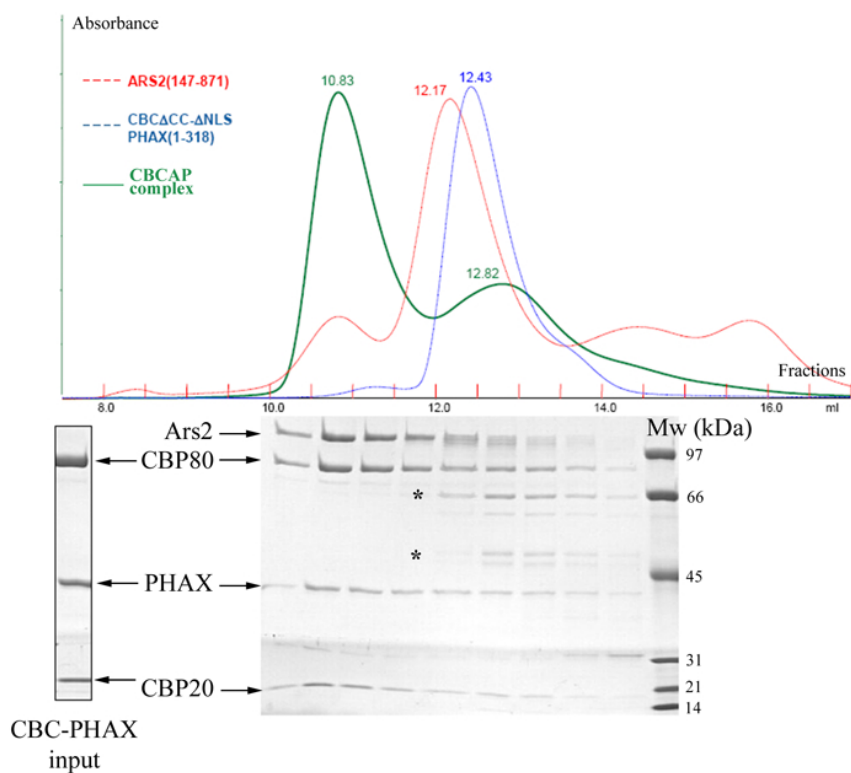
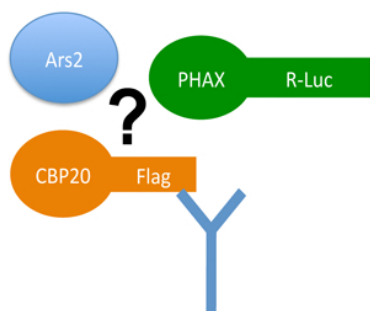
(A) SILAC proteomic analysis of Ars2 partners. U2OS cells stably expressing GFP-Ars2 were isotopically labelled and immuno-precipitated with anti-GFP antibodies, together with control cells bearing a different isotopic label. Pellets were analyzed by mass spectrometry. SILAC ratios of the identified proteins (enrichment over control) were plotted as a function of protein abundance (in Log10 scale). (B) Co-IPs demonstrate interaction between CLP1, and CBC or Ars2. Hela cells stably expressing Flag-Ars2 or Flag-CBP20 were transiently transfected with GFP-CLP1, extracts were treated or not with RNase, and immuno-precipitated with anti-GFP antibodies. Pellets were analyzed by western blot with anti-flag antibodies. (C) Same experiment as in B, except that a stable HeLa cell line expressing GFP-CLP1 from a BAC was used, and that anti-GFP pellets were analyzed with antibodies against endogenous Ars2 and CBP20, indicated with arrows. Star: contaminating band. (D) Depletions of CLP1 and PCF11 leads to read-through of U4, Hist1H3b and CENPW 3'-end processing sites. HeLa cells were transfected with the indicated plasmid and siRNAs, and Firefly luciferase activity was measured and normalized to that of Renilla luciferase. For each

plasmid, ratios of luciferase activity were additionally normalized to values of the control siRNA.

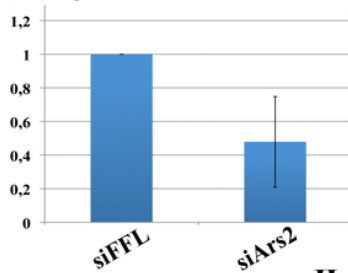
Figure 7: CBC and Ars2 preferentially stimulate cap-proximal sites

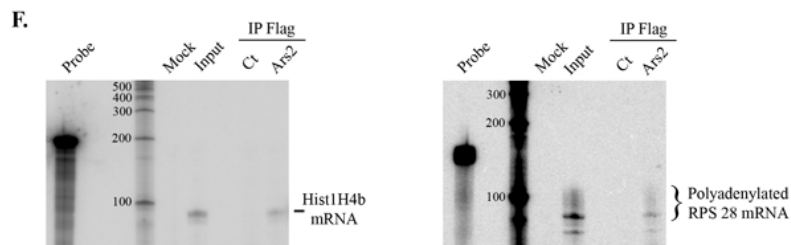
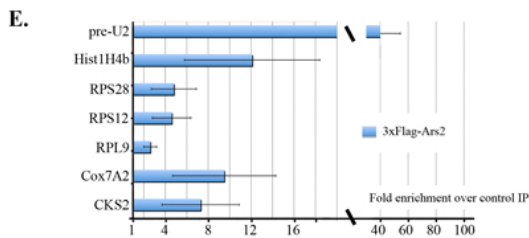
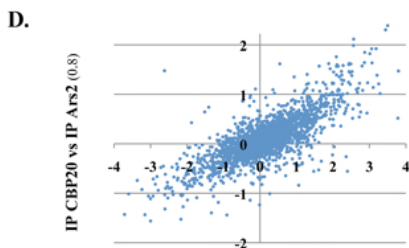
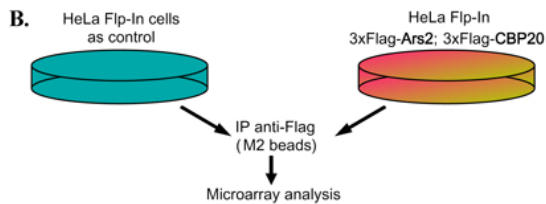
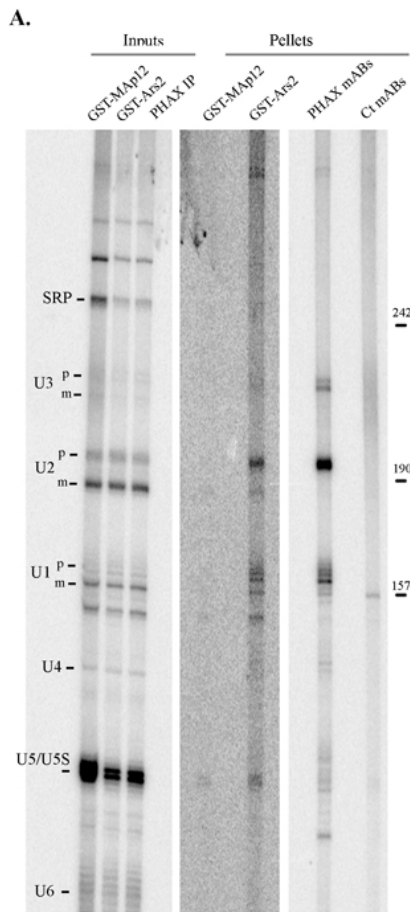
(A) Luciferase assay for readthrough of U4, Hist1H3B and CENPW processing site, with and without spacer sequences between the cap and the processing site. HeLa cells were transfected with the indicated plasmids and siRNA, and luciferase activity was measured and normalized as in Figure 6. Left: U4; middle: Hist1H3B; Right: CENPW. (B) Model depicting the coupling of snRNA processing and export by CBC and Ars2. (C) Model depicting the stimulation of cap-proximal processing by CBC and Ars2.

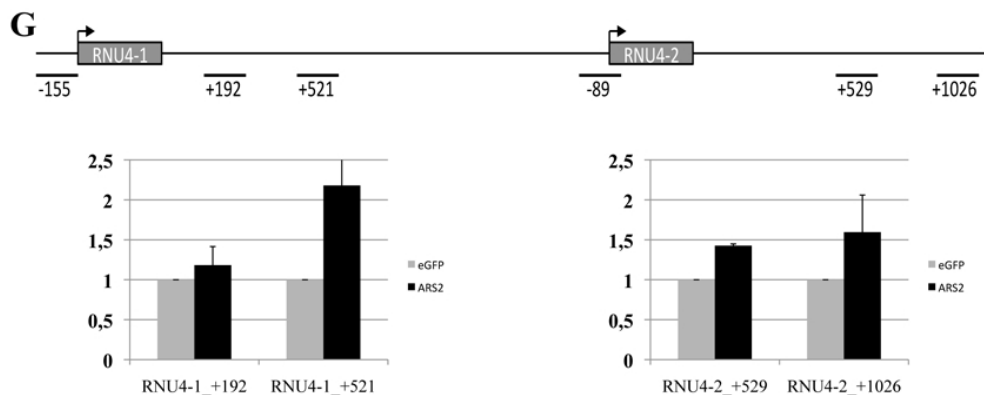
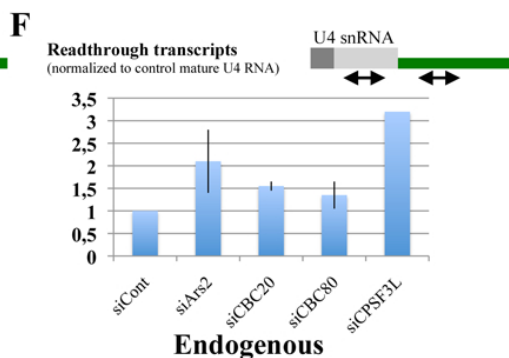
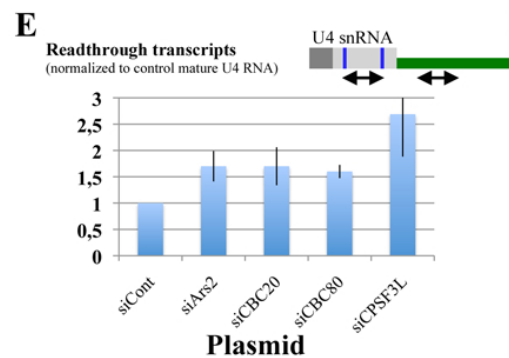
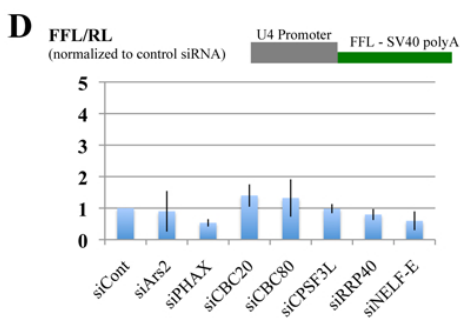
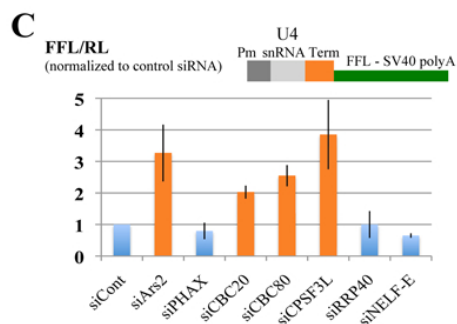
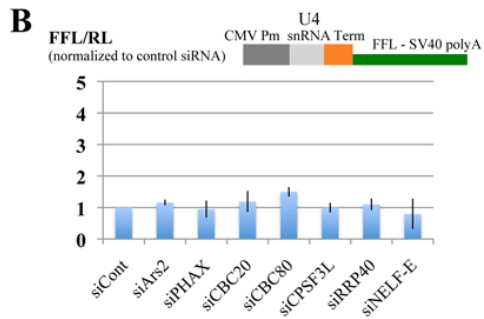
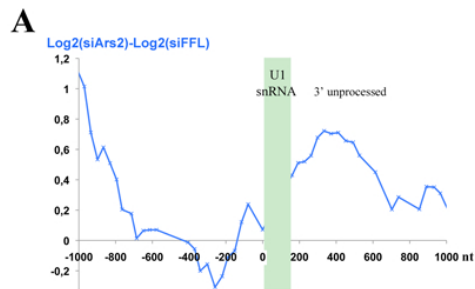


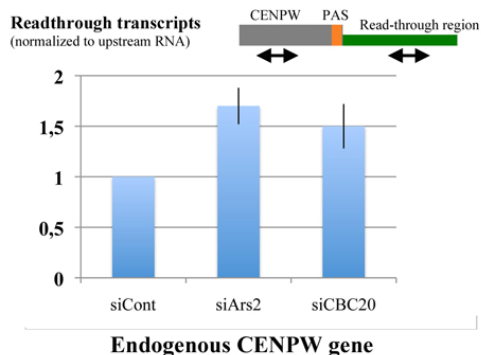
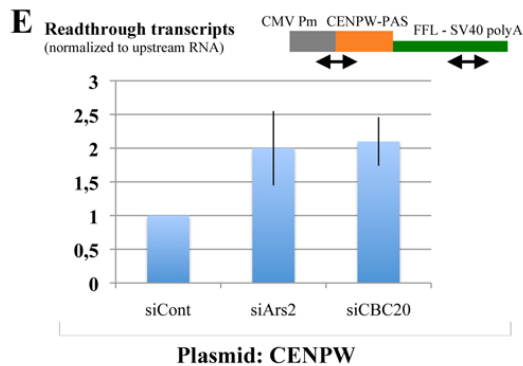
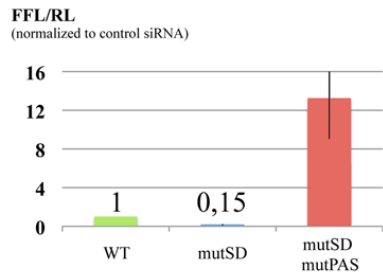
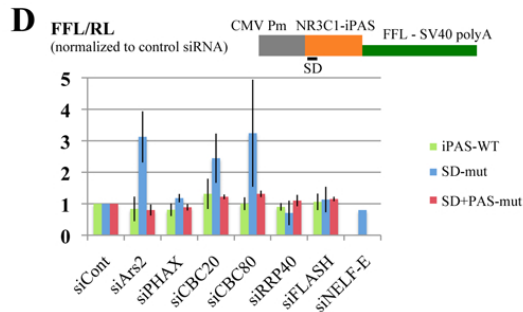
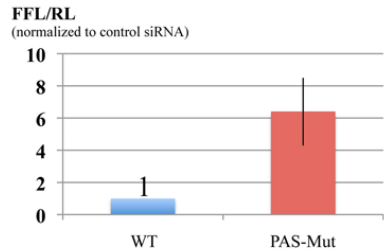
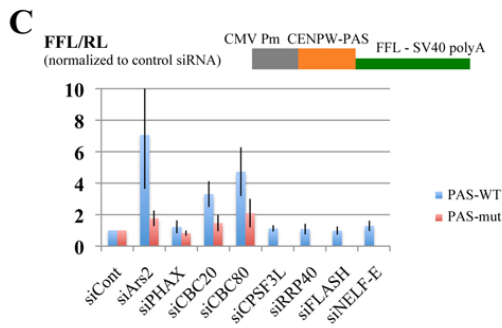
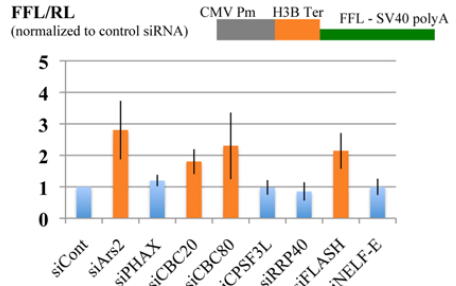
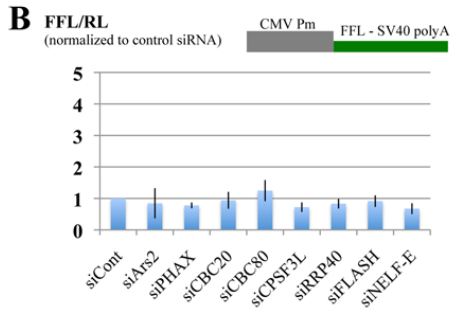
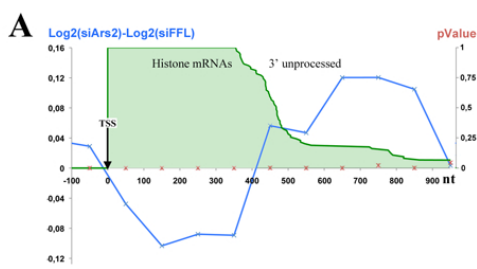
A.**B.****C.****Luciferase activity**

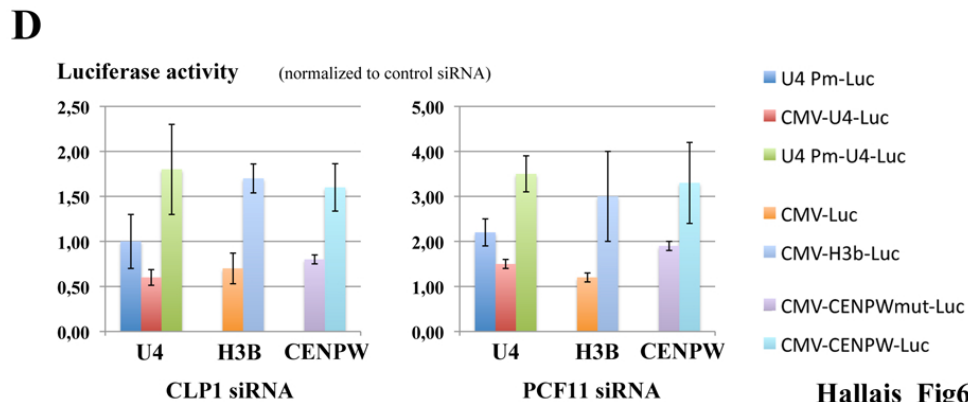
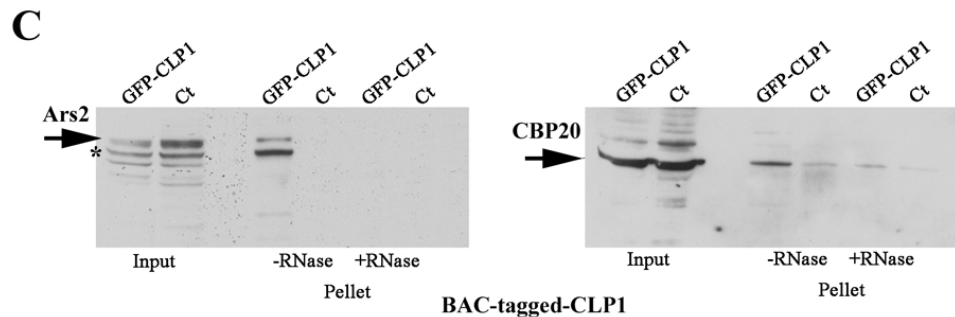
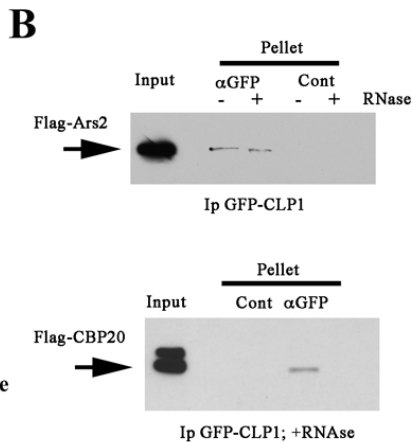
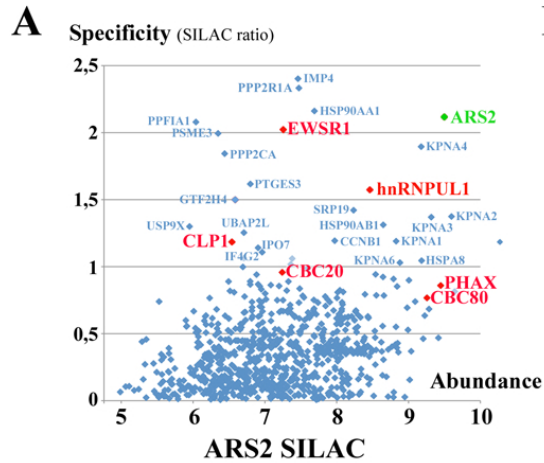
(beads/input; normalized to siFFL)

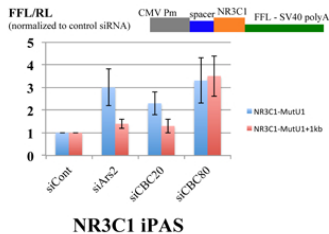
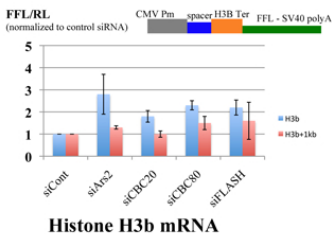
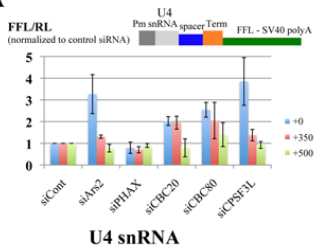
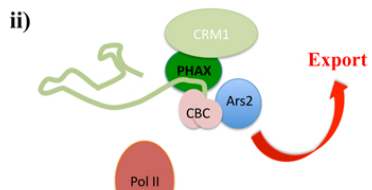
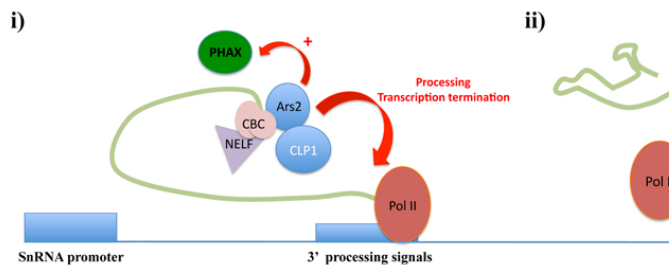










A**B****C**