

Emerging insights into the function and structure of the Integrator complex

Moritz M. Pfeleiderer  and Wojciech P. Galej 

European Molecular Biology Laboratory, Grenoble, France

ABSTRACT

The Integrator was originally discovered as a specialized 3'-end processing endonuclease complex required for maturation of RNA polymerase II (RNAPII)-dependent small nuclear RNAs (snRNAs). Since its discovery, Integrator's spectrum of substrates was significantly expanded to include non-polyadenylated long noncoding RNAs (lncRNA), enhancer RNAs (eRNAs), telomerase RNA (tertRNA), several Herpesvirus transcripts, and messenger RNAs (mRNAs). Recently emerging transcriptome-wide studies revealed an important role of the Integrator in protein-coding genes, where it contributes to gene expression regulation through promoter-proximal transcription attenuation. These new functional data are complemented by several structures of Integrator modules and higher-order complexes, providing mechanistic insights into Integrator-mediated processing events. In this work, we summarize recent progress in our understanding of the structure and function of the Integrator complex.

ARTICLE HISTORY

Received 21 December 2021

Revised 22 February 2022

Accepted 24 February 2022

KEYWORDS

Integrator; 3'-end processing; endonuclease; transcription attenuation; RNAPII; snRNA; cryo-EM

Introduction

The Integrator complex was originally discovered as the 3'-end processing machinery required for maturation of RNA polymerase II (RNAPII)-dependent small nuclear RNAs (snRNAs) [1]. Since its discovery, additional Integrator substrates have been identified, including non-polyadenylated long noncoding RNAs (lncRNA), enhancer RNAs (eRNAs), telomerase RNA (tertRNA), several Herpesvirus transcripts, and messenger RNAs (mRNAs) [1–5]. Today, Integrator emerges as a global regulator of RNAPII activity and one of the key factors modulating expression of protein-coding genes via transcription attenuation [6–8], a mechanism tightly linked to promoter-proximal pausing of the RNAPII and the presence of the corresponding transcription factors NELF and DSIF [9–14].

Integrator consists of 14 core subunits (INTS1–INTS14) [1,15], including β -CASP/MBL endonuclease INTS11, and several other accessory factors, like the recently reported protein phosphatase complex PP2A [16–18]. Together, both enzymatic activities are specifically targeted to RNAPII via a shared Integrator scaffold, allowing for the release of the nascent transcript and “resetting” of the phosphorylation state of the RNAPII C-terminal domain by removing phosphorylation marks.

Although the composition of Integrator was well-established over the years, until recently, there was very little information about its general architecture and molecular structure [16,17]. A recent influx of crystallographic and cryo-EM data provided structural information for almost all known Integrator subunits, including the Integrator cleavage module consisting of INTS4/9/11 [18], the INTS13/14 heterodimer [19], the Integrator-PP2A complex [20], and the Integrator-PP2A bound to RNAPII [11,14]. Together, these structures provide unprecedented mechanistic insights into the spatial arrangement of Integrator's subunits and their respective functions.

Here, we summarize the recently reported structures and discuss them in the context of our current understanding of Integrator's function.

Integrator acts on a diverse set of RNAPII-dependent transcripts

The Integrator complex was identified as one of the three cellular machineries performing 3'-end processing of RNAPII transcripts, alongside the Cleavage and Polyadenylation machinery (CPA) and the U7-dependent Histone pre-mRNA Processing Machinery (U7-HPM) [21–23]. The CPA complex is responsible for the 3'-end cleavage and polyadenylation of the vast

majority of protein-coding pre-mRNAs. An exception to that are replication-dependent histone transcripts, which are not polyadenylated, but instead processed by U7-HPM. A canonical maturation pathway for other non-polyadenylated RNAPII-dependent transcripts was unknown until Integrator was discovered [1,24].

In the context of snRNAs processing, it was postulated that specific signal sequences in the pre-snRNAs would be recognized by the Integrator for further substrate processing. Indeed, mammalian U1 and U2 snRNAs share a conserved GTTTN₀₋₃AAARNNAGA signal sequence (3'-box), located downstream from the mature 3'-end, which is required for their transcription termination and proper 3'-end processing [1,25,26]. Similar sequences were identified in several *Herpesvirus saimiri* Sm-class U RNA transcripts (HSURs), which were also shown to undergo Integrator-dependent processing (Figure 1) [3,5,27]. However, in this case, Integrator cleaves HSURs independently of RNAPII transcription, which stands in contrast to its activity on other RNA substrates [5]. Interestingly, in the absence of Integrator activity, cleavage and polyadenylation machinery and intrinsic properties of the sequences downstream of the snRNA coding sequences can contribute to transcription termination [28,29]. Moreover, some of the Integrator-dependent protein-coding genes contain 3'-box motifs near their transcription end sites (TESs) and were shown to have enriched RNAPII occupancy when analyzed by ChIP-seq experiments [13].

Although 3'-box sequences have been shown to be necessary for some of the Integrator-dependent 3'-end processing events, it remains unclear whether Integrator is directly involved in their recognition and whether it has any inherent sequence specificity. Biochemical studies with purified Integrator subunits could not detect any increased affinity to the substrates containing 3'-box motifs when compared to random sequences [18,19]. Recent evidence suggests that human snRNAs may exploit Integrator-sensitive gene regulation for cotranscriptional cleavage rather than recruitment to specific signal sequences [8,10].

In addition to snRNAs, Integrator was shown to be associated with several other coding and non-coding RNAPII-dependent transcription units,

including human lncRNAs, tertRNA, eRNAs, and mRNAs as well as piRNAs in *C. elegans* [2,4,8,30–32]. In all cases (including snRNAs), impaired Integrator's function results in the readthrough into neighboring transcription units and aberrant polyadenylation [8,10,30,33]. This effect is conserved among humans and different model organisms, including *C. elegans* [33], *D. melanogaster* [7], or planarian *S. mediterranea* [34].

The emerging picture suggests that all these diverse substrates could utilize a conserved mechanism of promoter-proximal transcription attenuation, which has different functional outcomes for different transcription units.

Mechanism of integrator-mediated transcription regulation

RNAPII undergoes a highly regulated transcriptional cycle, which includes recruitment of the polymerase to the genomic locus (Initiation), transcription elongation, release of the nascent transcript via endonucleolytic cleavage (3'-end processing), and separation of the polymerase from the genomic locus (Termination).

The C-terminal domain of the largest RNAPII subunit, Rbp1 (hereafter referred to as RNAPII^{CTD}), consists of multiple repeats of the heptad sequence YSPTSPS (52 repeats in humans) and is a subject to extensive phosphorylation throughout the transcription cycle [35,36]. Modification of the RNAPII^{CTD} phosphorylation state is a well-established mechanism in regulating RNAPII transcription activity, and many protein complexes, including Integrator, contain or associate with kinases and phosphatases for this purpose (e.g., transcription preinitiation complex: TFIIF; CPA: Ssu72; Mediator: CDK8).

Promoter-proximal pausing during early transcription elongation is a widely utilized regulatory mechanism in metazoans [37]. It takes place mainly within the first 40–60 nt after transcription initiation [37], when negative transcription elongation factors NELF and DSIF are recruited to the RNAPII and prevent them from further elongation [38]. Transitioning into the elongation state requires the activity of a CDK/cyclin kinase complex, P-TEFb, which phosphorylates RNAPII and pausing-related factors. This allows dissociation of

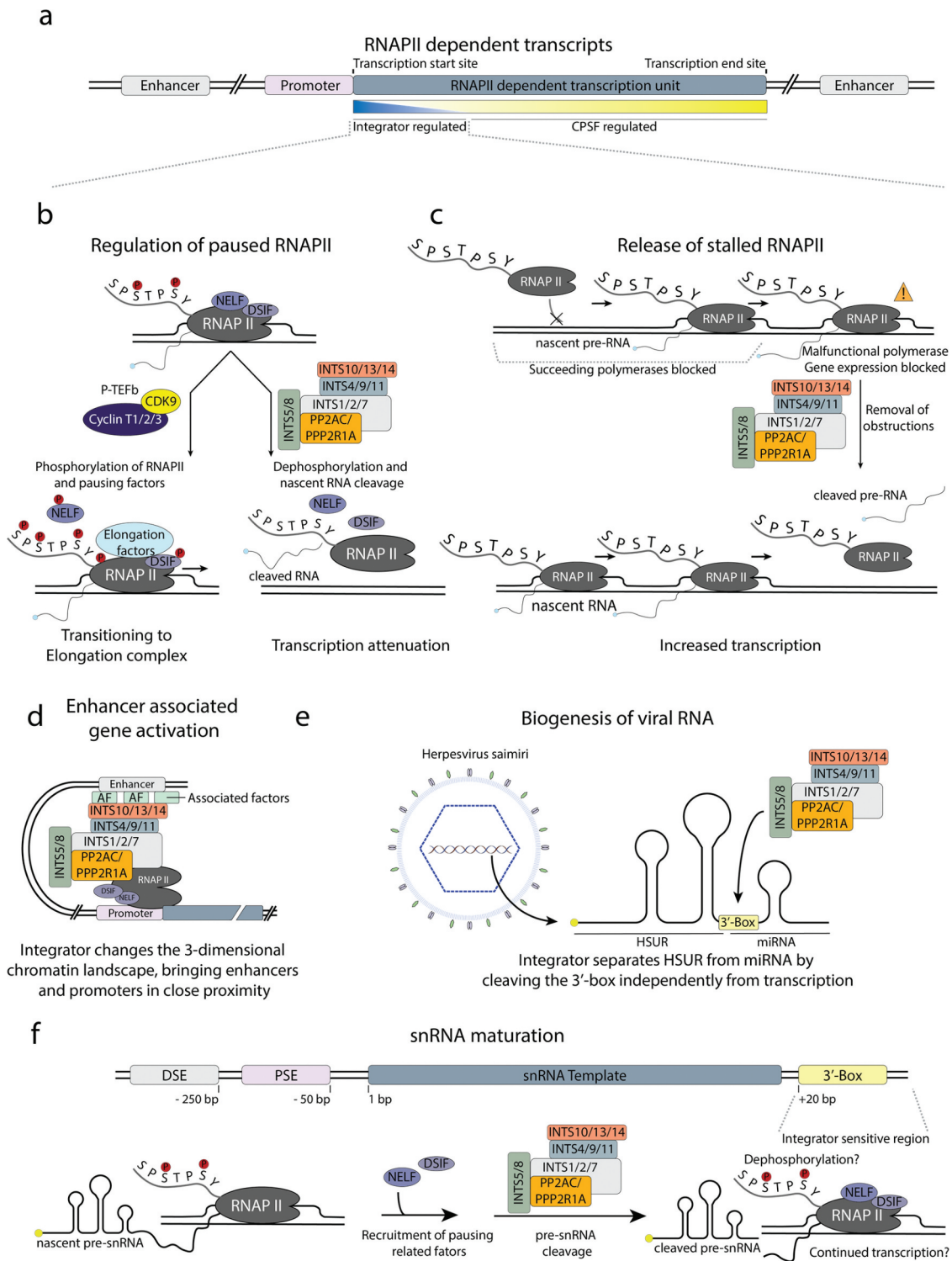


Figure 1. Different modes of Integrator function.

a Schematic representation of an RNAPII-dependent transcriptional unit with the promoter proximal regions sensitive to Integrator regulation. **b** Possible fates of paused RNAPII. Recruitment of the Integrator results in cleavage of the nascent RNA and dephosphorylation of the RNAPII, leading to transcription attenuation. Recruitment of P-TEFb promotes transitioning into a transcription elongation complex. **c** Integrator-mediated removal of stalled RNAPII frees the target gene from obstructions and facilitates further cycles of transcription. **d** Integrator-mediated transcription regulation by affecting the 3-dimensional chromatin landscape, bringing enhancer and promoter elements into close proximity to induce transcription. **e** Processing of *Herpesvirus saimiri* RNAs by the host Integrator is independent of RNAPII. **f** Cartoon representation of an snRNA gene with associated regulatory elements and the indicated region sensitive to Integrator processing. Pausing-related factors are recruited and facilitate Integrator-mediated release of the pre-snRNA.

NELF and recruitment of other factors to form the Elongation complex [37,39].

Mapping of the chromatin occupancy of Integrator subunits revealed a preference for promoter proximal regions, within the first 3000 bp of the transcription start site [7–9]. In line with this observation, Integrator subunits were shown to interact with the negative transcription elongation factors NELF and DSIF, which are necessary for Integrator recruitment to RNAPII on protein-coding and snRNA genes [9,10,12,13]. Coupling to RNAPII is further strengthened by recognition of specific phosphorylation marks in RNAPII^{CTD}, in particular, Ser2/Ser7 [40] and Tyr1 [41].

Recruitment of Integrator to the paused RNAPII has two major consequences. First, it allows the nuclease, INTS11 to cleave and release nascent RNA. This results in small RNA fragments, which are rapidly degraded by the exosome, or in functional noncoding RNA products [6,8,11]. In addition, cleavage of nascent transcripts exposes uncapped 5'-ends of the newly synthesized RNA, which are targeted by the exonuclease Xrn2, leading to transcription termination [11,42]. Second, Integrator mediates recruitment of protein phosphatase 2 (PP2A) to the stalled polymerase complex, which allows dephosphorylation of the RNAPII^{CTD}. Cross-linking data shows that Integrator binds exclusively to the C-terminal half of the RNAPII^{CTD}, but it is unclear whether the entire CTD is dephosphorylated or only particular regions [11]. Integrator-associated phosphatase activity is particularly pronounced for residues Ser2, Ser5, and Ser7 [20,43]. By removing phosphorylation marks from the RNAPII^{CTD} and the associated factors (i.e., NELF and DSIF), the Integrator-PP2A complex facilitates turnover of RNAPII and transcription attenuation [44,45]. Dynamic transcription termination and reinitiation of RNAPII at gene promoters have been shown to play an important role in promoter-proximal pausing by competing with transitioning into the transcription elongation phase [46–48]. This concept is well established across different species and provides an important layer of regulation of gene expression [49–53].

Both enzymatic functions of Integrator contribute to the same outcome, but several studies suggest that their activity is not coupled, but they work independent of each other, e.g., depletion of the nuclease module has no effect on the phosphatase function [6,20,54]. Furthermore, the phosphatase activity of

Integrator seems to be dispensable for snRNA processing, as indicated by reporter assays and ChIPseq data, suggesting that snRNA processing and transcription attenuation are at least partially distinct processes [6,20,55,56].

By aborting transcription cycles, Integrator acts as a global attenuator of gene expression [7,57]. Accordingly, RNAseq experiments identify a broad range of genes that are upregulated upon Integrator depletion in *Drosophila* and in mammalian cells [6,7,57]. For a subset of genes, however, the presence of Integrator boosts transcription [6,7,57]. The first class (downregulated by Integrator) deals with genes prone to promoter-proximal pausing of RNAPII. This event takes place in only a subset of protein-coding genes, in particular, immediate early genes, even though the exact mechanisms that define pausing remain unclear [37]. Current models suggest that those genes are always transcribed but rely on external signals to progress from pausing into the elongation phase (i.e., phosphorylation by P-TEFb) and that Integrator aborts transcription at this stage (Figure 1b) [6–9,43].

The second class of genes (upregulated by Integrator) are not subjects to active Integrator-mediated downregulation, but concern the situation, when RNAPII will occasionally become stalled in a nonproductive state. These nonfunctional, stalled polymerases are also recognized by Integrator and removed, thus freeing the gene from obstructions for further transcription cycles (Figure 1c) [7,57].

By recognizing paused RNAPII, Integrator and P-TEFb are direct antagonists and compete for the same substrate of promoter proximally paused RNAPII [55]. Paused polymerase, which is recognized by P-TEFb, is phosphorylated and may transition to a productive elongation state, whereas recruitment of Integrator leads to abortive transcription (Figure 1) [55]. It is possible that the PP2-Integrator – P-TEFb – axis evolved as a phosphatase/kinase switch to fine tune transcription. Loss of both regulators results in severely decreased transcription activity as neither pause/release nor rescue of the stalled polymerase are possible [55].

In addition to its role in the promoter-proximal transcription attenuation, Integrator has been implicated in additional function near transcription end sites (TES) [13]. Under hyperosmotic stress conditions, association of the Integrator components with

RNAPII is impaired, resulting in hundreds of downstream-of-gene (DoG) readthrough transcripts generated from upregulated genes [58]. Depletion of the Integrator catalytic subunit, INTS11, recapitulates this effect [58] and leads to aberrant transcription termination and 3'-end processing of a subset of protein-coding genes, even in the presence of the canonical CPA [59]. This effect is most prominent near alternative polyadenylation sites and for sequences with potential to form secondary structures [59]. Similarly, it could be shown that loss of the elongation factor SPT6 leads to readthrough in Integrator-dependent lncRNA [30]. The mechanism of the Integrator recruitment and activation at the end of genes remains to be investigated.

Composition of the integrator complex

Integrator was originally discovered as a 12-subunit protein complex that associates with RNAPII, while searching for interactors of DSS1 (Deleted in split hand/ split foot 1), a protein that interacts directly with BRCA2 and the proteasome [1]. Two additional subunits INTS13 and INTS14 were identified in an RNAi screen performed in *D. melanogaster* [15]. Finally, Integrator was shown to recruit Protein Phosphatase 2 (PP2A), which is associated with many different protein complexes [20,60,61].

Experimental evidence suggests that the core Integrator complex can be split into multiple stable modules including the cleavage module (INTS4/9/11) [19,50], the shoulder module (INTS5/8) [19,20], and a ternary complex of INTS10/13/14 [18,45].

Several additional proteins have been associated with Integrator but are not considered part of the core complex [62]. These include the DNA damage repair factors SOSSB1/2 and SOSSC [63–65] as well as EGR1/2 and NAB2 [54]. Interestingly, the function of these associated proteins seems to differ from the canonical nuclease/phosphatase activity of the Integrator. While SOSSB1/2 and SOSSC interact with INTS3 and INTS6, they were not shown to exist in the context of the remaining twelve Integrator subunits, suggesting that the Integrator complex may consist of sometimes mutually exclusive subunits [16,66,67]. Considering that the Integrator is a highly modular protein machinery, it was postulated that it may function in multiple different compositional and conformational states and some modules may be

required for additional tasks or work outside of the canonical Integrator function [15,56].

The structure of the integrator complex

Cleavage module – INTS4/9/11

Among all Integrator subcomplexes, the cleavage module is structurally best characterized [11,18,20,68]. While most Integrator subunits are mainly α -helical with very few predictable domain features, INTS9/INTS11 adopt MBL/ β -CASP folds with remarkable similarity to their corresponding orthologs CPSF100/CPSF73 [69]. Both nucleases arrange head-to-head with a pseudo-two-fold symmetry running along their interfaces [18,20]. Their C-terminal domains (CTDs) are tightly intertwined and form an extensive interface between the two proteins, while the nuclease domains remain more loosely associated [18]. Interestingly, the CTDs of INTS9 and INTS11 can be divided into two subdomains, both of which form compact composite interfaces with one another, but only the very C-terminal CTD2 is required for dimerization [17,18]. The other composite domain CTD1 forms only upon successful dimerization of CTD2. Proper formation of this quaternary structure is required for the recruitment of INTS4, which, in turn, integrates the nuclease dimer into the fully assembled Integrator complex [18,20]. INTS4 consists of HEAT-repeats stretching along the MBL domain of INTS9 and subsequently wedging between the interface of both nucleases [18]. The C-terminal region of INTS4 forms a β -sandwich, which sits on top of the β -CASP domain of INTS9. Together, INTS4/9/11 forms a highly positively charged cavity, which serves as a conserved binding site for inositol hexaphosphate (IP6) [68]. Mutations close to the IP6-binding site cause nuclease malfunctions in reporter assays; however, it is unknown whether IP6 contributes to Integrator activity or has only a scaffolding function [18,68].

INTS10/13/14 module

Integrator subunits 10, 13, and 14 form a stable heterotrimer, which is further associated with the cleavage module [18,19,54]. Reporter assays could show that INTS13/14 is crucial for snRNA processing in *Drosophila* [15]; however, in human cells,

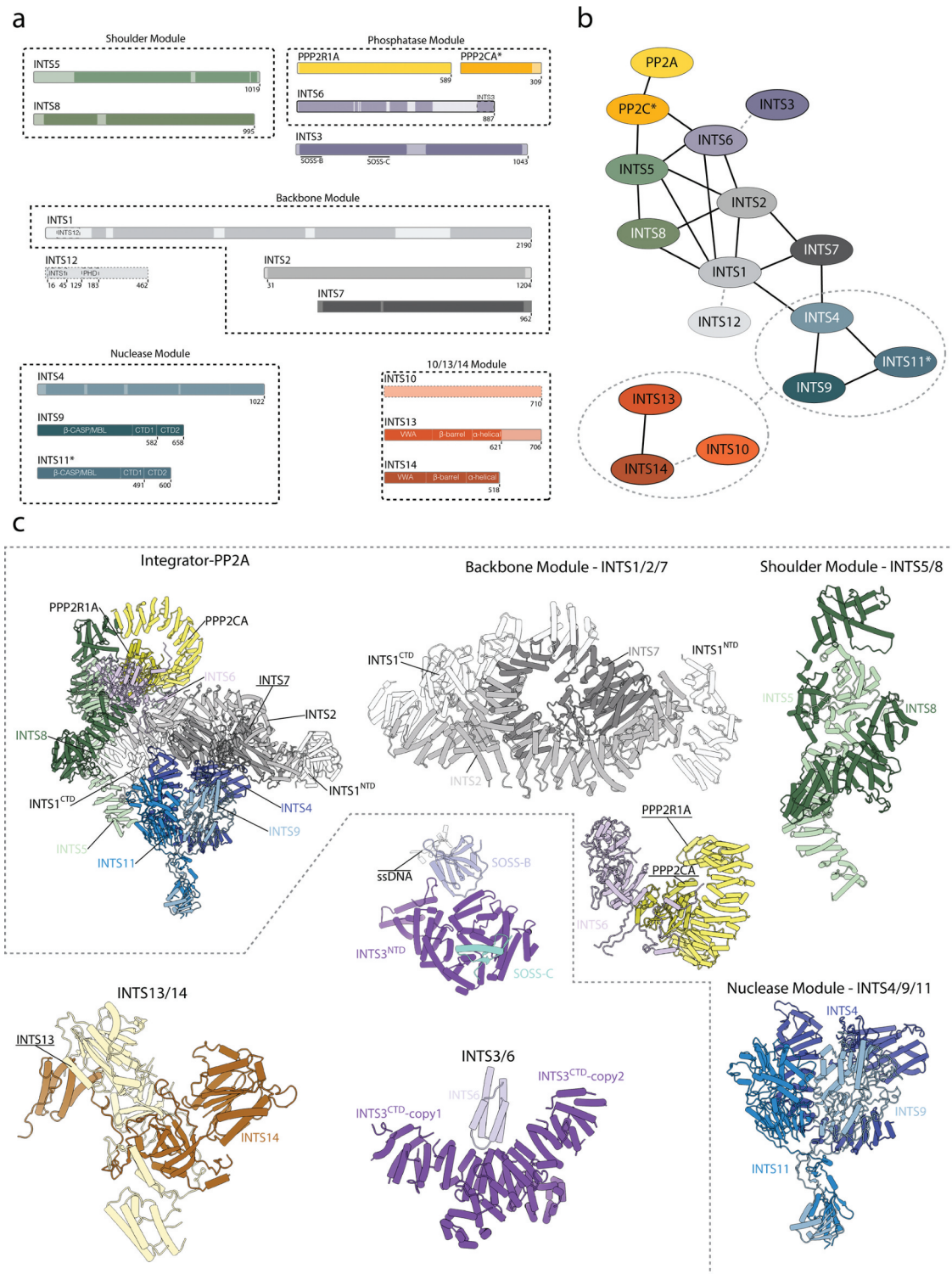


Figure 2. Structure of the integrator-PP2A complex.

a) Sequence bar representation of all subunits forming the Integrator-PP2A complex. Regions with structural information available are opaque, transparent regions, and dotted lines correspond to sequence regions, which are not resolved in the reported cryo-EM/X-ray experiments. No structural information is available for INTS10 and INTS12. **b)** Diagram showing the interaction network of the different Integrator subunits. Dark lines correspond to interfaces confirmed by structural data, and dotted lines are contacts inferred from biochemical data. **c)** Structure gallery showing all known Integrator structures. Figures based on 7CUN (Integrator-PPA complex) [20], 6S11 (INTS13/14) [19], 7BV7 (INTS3/6) [67], and 4OWW (INTS3/SOSS-A/SOSS-B) [16].

none of the three subunits appear to be critical for endonuclease activity *in vivo* [4,12,41]. Although the precise function of this module remains unknown, it was shown to play a role in activating poised enhancers during cell differentiation [54]. Studies capturing the spatial arrangement of chromatin showed that depletion of INTS13 correlates with decreased enhancer/promoter interactions [2,54]. INTS13 was also described as a factor functioning independent of the canonical Integrator

and interacting with EGR1/2 and NAB2, which promote interactions with the enhancer elements [54].

Structure prediction suggests that INTS10 is a mostly α -helical protein and a crystal structure of INTS13/14 shows both proteins to form a pseudo-symmetric heterodimer with a complex interface running along all domains of both proteins [19]. Both proteins consist of a N-terminal VWA domain followed by a β -barrel, a linker, and

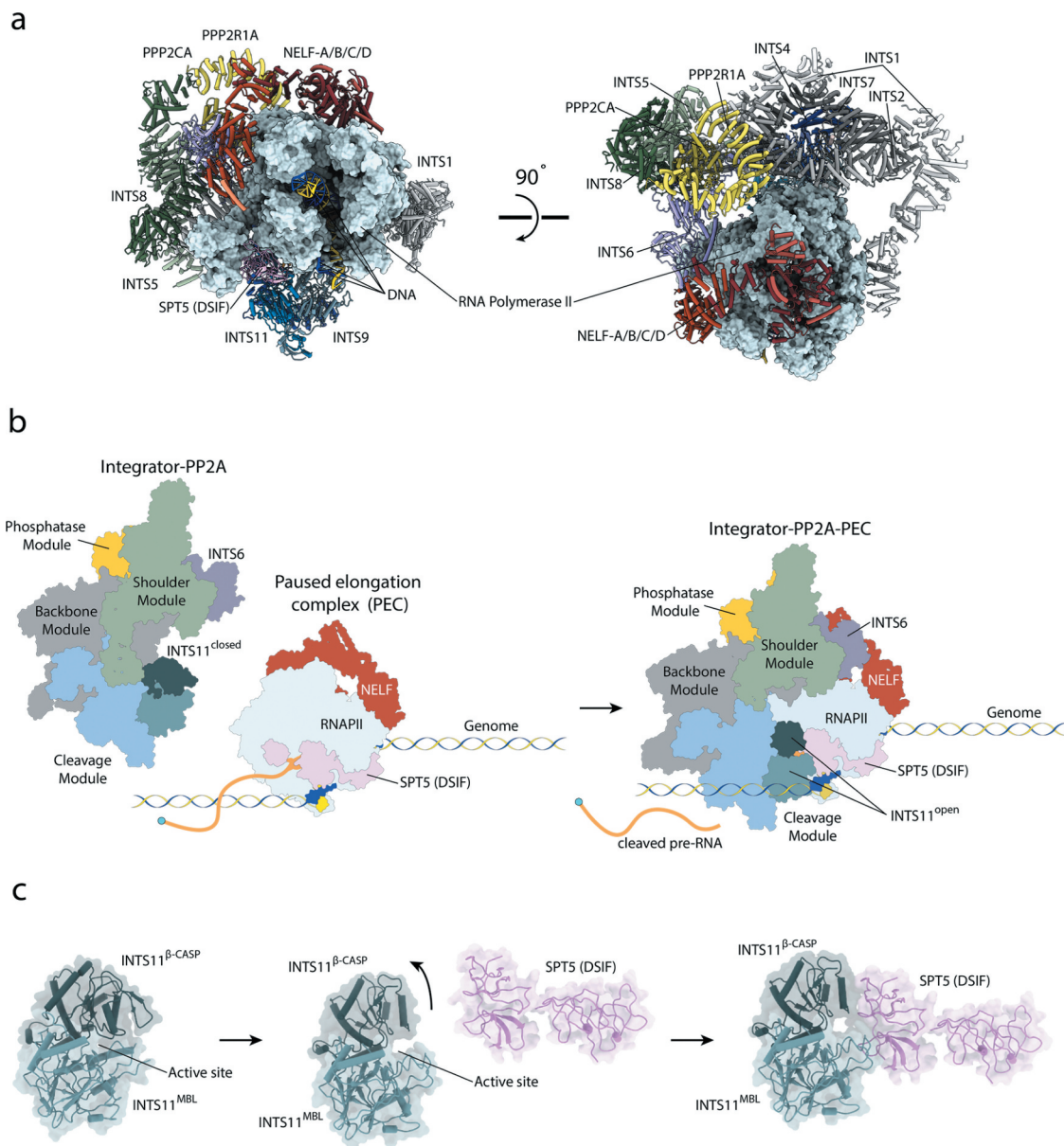


Figure 3. Integrator association with RNAPII and activation of the endonuclease.

a) Recruitment of Integrator to RNAPII. RNAPII is shown in surface representation. Integrator and pausing factors NELF and DSIF are depicted as colored cartoons [11]. **b)** Mechanism of the Integrator recruitment to paused RNAPII and activation of the INTS11 endonuclease. **c)** A close-up view of INTS11 and its interaction with SPT5 shown in the same orientation as in b). Figures are based on PDB coordinates: 7BFP [18] and 7PKS [11].

an α -helical region, which in the case of INTS13 is further extended by a 130 residue long, flexible CTD. This region encompasses a cleavage module-binding motif (CMBM), responsible for the interaction with INTS4/9/11 [19]. The interaction of INTS10 with INTS13/14 was mapped to a MIDAS pocket in INTS14 [19]. The exact function of the INTS10/13/14 module remains elusive, but based on the structural similarity to the DNA-binding protein Ku70/80, it was proposed to be required for RNA binding [19]. However, RNA binding assays could not detect specificity or high affinity for model substrates, as observed for modules of the CPA complex [18,19,70]. To date, none of the Integrator structures provide any information about the exact position of the INTS10/13/14 module relative to the remaining subunits [11,20].

The integrator – PP2A complex

A cryo-EM structure of the fully assembled, recombinant Integrator-PP2A complex (INTAC) revealed an overall organization of the Integrator subunits and their interactions with the accessory PP2A phosphatase complex (Figure 2) [20].

The core of the Integrator complex is composed of the subunits INTS1/2/7 forming the “backbone module”, which forms the scaffold for the nuclease (INTS4/9/11) and phosphatase modules (PPP2CA, PPP2R1A). Both enzymatic activities are located at opposing ends of the structure, separated by approximately 150 Å and seem to function independent of each other (Figure 2) [6,20].

The nuclease dimer INTS9/11 is recruited via INTS4 to INTS7 of the backbone module. INTS5/8 form the “shoulder module”, which is required for recruitment of the PP2A phosphatase to the Integrator scaffold [20,43]. PP2A family phosphatases are highly diverse protein complexes typically consisting of a catalytic subunit, a scaffolding subunit, and a regulatory subunit [61]. However, only the catalytic and scaffolding subunits are found in the Integrator [20]. Furthermore, Integrator seems to be specific for the scaffolding protein PPP2R1A and the catalytic subunit PPP2CA, whereas their homologues were detected at lower quantities [20]. PP2A is further stabilized by interactions with INTS6, which mediates contacts to INTS2 of the

scaffolding module and to the shoulder module. The interaction of INTS6 with the shoulder module alone seems to be rather weak as it could not be copurified together with INTS5/8 [18]. In addition, INTS6 is a direct binding partner of INTS3 and previous interaction studies map their interface to their respective C-terminal regions [66,67,71]. Only the N-terminal region of INTS6 is ordered, and in line with this observation, INTS3 remains disordered in all available Integrator-PP2A structures [71]. INTS10/13/14 and INTS12, although present in the sample preparation, remain flexible and were not visualized in the Integrator-PP2A structure.

Integrator – PP2A – PEC

Integrator is recruited specifically to RNAPII bound to NELF and DSIF (Paused Elongation Complex – PEC) [7,9,12,13,43]. The structure of the PEC alone was reported previously, and its general architecture is identical to the Integrator-bound PEC [11,38] (Figure 3). However, parts of DSIF become disordered upon the recruitment of Integrator-PP2A [11].

The Integrator wraps around the polymerase on the opposite site of the cleft with the RNAPII active site and the DNA/RNA duplex (Figure 3a). In order to accommodate RNAPII, Integrator-PP2A undergoes several subtle conformational changes, which allow it to engage with the polymerase via four interfaces [11]. The main interface is formed between the N-terminal region of INTS1 (around residues 350–600) and RBP2. The second, smaller interface depends on INTS7 binding to RBP3. Two other interactions involve transcription pausing factors, where the nuclease INTS11 binds the KOWx-4 domains of SPT5 and INTS6 to NELF-B [11].

The MBL/ β -CASP domain of INTS11 is a bona fide nuclease, but remains in an inactive state until its recruitment to paused RNAPII [11,18]. Two recent cryo-EM structures [11,14] revealed that the KOWx-4 domain of SPT5 (DSIF) pushes against the β -CASP lid of INTS11 (Figure 3b-c), opening it and allowing a substrate to engage with the active center [11]. The RNA exiting the active site of RNAPII is partially protected by the KOWx-4 domain of

SPT5 and is guided directly into the active site of INTS11. The distance of the active site of RNAPII to the active site of INTS11 encompasses only 20 nt [11]. Thus, the length of the released RNA depends mainly on the position of the RNAPII pause site during Integrator recruitment, in agreement with the sizes of the short, Integrator-dependent unstable transcription products [6–8,11].

Binding of RNAPII^{CTD} to the Integrator is mediated via a composite interface of INTS4/2/7, which forms a cavity that accommodates the RNAPII^{CTD} with sidechain-specific contacts. Locations of additional peptides are reported by Zheng et al. with increasing proximity to the active site of PP2A, suggesting a preferred path of the RNAPII^{CTD} within the Integrator complex [14]. However, no density could be observed close to the active site of PP2AC [11]. It is interesting to note that, while the presence of pausing factors is required for nuclease activation, RNAPII alone could be sufficient to recruit Integrator and execute its phosphatase activity.

Recruitment of Integrator is mutually exclusive with the recruitment of the Mediator complex and other factors forming the transcription preinitiation complex (PIC). Similarly, other transcription factors like PAF1C or SPT6 are not compatible with the presence of the Integrator [11,14,72].

INTS3 and DNA damage repair complexes

INTS3 is a well-established subunit of a DNA damage repair complex, SOSS (sensor of ssDNA), that acts autonomously from Integrator. N-terminal α -helical repeats of INTS3 form the scaffold for the two interactors SOSSB1/2 (alternatively hSSB1 – human ssDNA binding protein 1/2) and C9ORF80 (SOSS-C, MISE – minute INTS3/hSSB1-associated element, INIP – INTS3 interacting protein) [63–65]. SOSS was described before as a major factor in maintaining genome stability, aiding detection of DNA lesions and recruiting factors downstream of the DNA damage repair pathway [63,71,73]. INTS3 interaction with SOSS-B and SOSS-C is sensitive to UV-light exposure, which as well functions as a signal for transport into the nucleus and localization to chromatin [74]. A crystal structure of the INTS3^{NTD} together

with SOSS-B, SOSS-C, and ssDNA could show that SOSS-B1 adopts an oligonucleotide binding fold, interacting with the bases of a 10 nt long ssDNA [16]. The function of SOSS-C remains elusive as it is dispensable for DNA binding [16]. Biochemical studies and a crystal structure of the INTS3^{CTD} identified an additional, putative nucleic acid binding site with a μ M affinity and preference for RNA [66,75].

In addition, a direct interaction of INTS3 and INTS6 was reported [63]. Immunoprecipitation experiments and a crystal structure could show that the CTD of INTS3 mediates interaction with INTS6 and may be involved in multimerization [66,67]. Based on these interaction studies, it is likely that the heterotrimeric complex of INTS3/INTS6/hSSB1 functions independent of the Integrator [63]. However, it cannot be excluded that this module is only weakly tethered to the fully assembled Integrator and is only required under certain conditions.

Common design principles of the 3'-end processing machineries

3'-end processing is a crucial step in gene expression, and factors involved in it are often highly conserved and fine-tuned to recognize their cognate substrates. Among them, the CPA complex is conserved from yeast to humans [22], while the U7-HPM and Integrator are restricted to metazoans [76].

Nonetheless, several features are common to all three 3'-end processing machineries like the catalytic core composed of an active and inactive MBL/ β -CASP nucleases [18,77]. CPA and U7-HPM share the same catalytic core composed of CPSF73 and its inactive partner CPSF100 [22], while Integrator uses the homologous INTS11 and inactive INTS9, respectively. Integrator does not share any subunits with the other two machineries, nor does it have any homologous subunits, besides the INTS9/11 nuclease heterodimer [1]. Both, CPSF100/73 and INTS9/11, nuclease heterodimers are arranged in a head-to-head orientation with a pseudo two-fold symmetry axis running along their dimerization interface. Stable interaction is mainly driven by their respective CTDs,

which are tightly intertwined and form two composite domains, CTD1 and CTD2.

Noteworthy, the inactive nucleases, CPSF100 or INTS9, contain noncanonical insertions in their respective pseudonuclease domains. In the case of INTS9, the NAD (Nine Accessory Domain) is composed of two insertions in the MBL domain, which forms a compact domain [18]. In contrast, the CPSF100 insertion is mostly unstructured and is located in the β -CASP lid [78]. However, it contains a short signal peptide (PIM – Polyadenylation Specificity Factor Interacting Motif), which tethers it to the polyadenylation specificity factor, the protein complex that recognizes polyadenylation signal sequences in the nascent pre-mRNA [78].

The nuclease catalytic subunits, INTS11 and CPSF73, are kept in inactivated states and require additional factors to achieve catalytic competence. As described earlier, Integrator undergoes activation through association with promoter-proximally paused RNAPII, where negative transcription elongation factor SPT5 pushes into the β -CASP lid of INTS11, displacing it and making the active center accessible [11,14]. U7-HPM utilizes a different mechanism, in which a protein from the LSm ring (i.e., Lsm10) inserts between the MBL and β -CASP domain of CPSF73 and allows the RNA substrate to engage with the active site [78]. The fact that the LSm ring associates with the U7 snRNA, which is responsible for histone pre-mRNA recognition, links the catalytic activation with the substrate recognition. However, Lsm10 is not present in the CPA, implying that it must utilize a different activation mechanism. Recent studies show that Rbbp6 or its yeast homolog Mpe1 is required for CPA catalytic activation [79–82]. To date, no structural information is available for the activated nuclease CPSF73 in the context of the CPA, and thus, the exact mechanism of the active site opening mechanism remains elusive.

It is interesting to note that both CPA and Integrator incorporate nuclease and phosphatase activities, which is not the case for the U7-HPM. This is consistent with the fact that Integrator and CPA interact directly with the RNAPII^{CTD} to regulate its phosphorylation states, while a similar interaction seems to be missing in U7-HPM-

mediated 3'-end processing. In the CPA, Symplekin recruits the phosphatase Ssu72 to the complex. Although Symplekin is shared between CPA and U7-HPM, the Ssu72 binding interface is unavailable in the U7-HPM, which exemplifies the specialization of both machineries [77,83].

The highly conserved architecture of the cleavage modules suggests a common evolutionary origin of the core of all 3'-end processing machineries. Additional subunits, which are not shared between different machineries, were likely recruited to each complex during adaptation to specific tasks.

Functionally, the main difference in Integrator and other 3'-end processing factors is that Integrator nuclease and phosphatase activity has a mainly abortive impact on transcription, while CPA and the U7-HPM are key factors required for the maturation of transcripts. Exceptions like snRNA exploit Integrator-sensitive regulation to achieve different functional outcomes [8].

Integrator involvement in development and cell differentiation

Multiple studies show that impaired Integrator function disrupts the fine-tuned equilibrium of specific transcripts required for cell differentiation, resulting in disease relevant phenotypes on a cellular level and in whole organisms. These effects are conserved among different species [84].

In *Drosophila*, Integrator has been implicated in the differentiation process of neuronal progenitor cells [85]. The study identified more than 1400 target genes regulated by INTS5, including transcription factors such as CycE and Notch signaling components, previously proposed to be involved in cell differentiation [85].

Similar results were reproduced in Zebrafish, where Integrator was shown to be crucial for embryonic development [86]. Here, INTS6 was shown to be a central factor in regulating dorso-ventral patterning by disrupting expression of crucial signaling factors such as BMP ligands and mediators of the Wnt signaling pathway [86].

In mouse models, Integrator was shown to regulate the differentiation of adipocytes [87]. Depletion of Integrator subunits resulted in altered

mRNA levels. In addition, the authors were able to correlate INTS expression with the process of differentiation pointing at the Integrator as a key factor in its regulation. Another study was able to show that Integrator is required for hematopoiesis [88]. Homozygous deletion of most INTS in mice is lethal in preweaning states, further highlighting the importance of Integrator in development [89].

Integrator malfunctions and genetic disorders in humans

The Integrator complex is involved in regulation of transcription homeostasis in human cells, in particular, in development and under stress conditions, and therefore, it is not surprising that its malfunctions are associated with genetic disorders and cancer [6,55,84].

In 2017, first studies emerged linking neurodegenerative disorders to mutations in INTS1 and INTS8 in several individuals from different families [90]. In all cases, biallelic Integrator mutations resulted in severe phenotypes with strongly impaired cognitive abilities [90–92]. For INTS8, one allele produced an unstable RNA due to a missense mutation, whereas the second allele contained a 3-residue deletion, which was shown to be the predominant protein in the patients. Comparing Integrator assembly in the mutated or WT variant of INTS8 showed a significant reduction in the enrichment of other INTS in human tissue culture [90]. Consistent with biochemical data, patients lacking functional INTS8 showed changes in splicing patterns and several transcripts were significantly up- or downregulated compared to healthy individuals [7,8,90]. Similarly, the mutation of INTS1 could be associated with a strong reduction of mRNA levels in skin fibroblasts [90].

A detailed study evaluating the genome and transcriptome of different cancer types revealed that all 14 Integrator subunits are frequently mutated in cancer cells [93] and dysregulate various cellular functions [94,95]. However, as several INTS are part of the genome maintenance machinery, possibly outside of the canonical Integrator function, it is difficult to establish clear causative links between those observations [96–98].

By regulating a large subset of RNAPII-dependent transcripts, Integrator is in principle capable of affecting various aspects of cell function, explaining its involvement in sometimes seemingly unrelated processes.

Open questions and future perspectives

Since Integrator was discovered, our understanding of its cellular function has undergone a paradigm shift. It was originally thought to be a specialized 3'-end processing machinery required for a very specific set of substrates, but recently emerged as a global regulator of the RNAPII activity [1,8]. Functional data are backed up by a gallery of different structures providing detailed insights into the exact spatial arrangement of the different subunits, their recruitment to RNAs, and activation of the nuclease subunit [11,18–20]. Nonetheless, several aspects of Integrator function remain elusive.

First, it is still unknown how Integrator recruitment to paused RNAPII is regulated. In general, P-TEFb and Integrator are present together in cells and compete for access to the stalled RNAPII. The modulation of this competition, in particular upon activation of signaling pathways and across different cell types, remains to be investigated.

Second, our current structural understanding of Integrator activity allows for assignment of functions to almost all Integrator subunits with exception of INTS3, INTS12, and INTS10/13/14.

INTS3 is also part of a DNA damage repair complex, but its functions here or in transcription attenuation remain elusive. It is unclear how the INTS3/6/SOSSB/SOSSC module relates to the Integrator complex and what is the function of INTS3 without its binding partners SOSSB and SOSSC.

INTS12 is a largely unstructured protein, which supports the stability of INTS1 and may be required for recruitment of RNAPII [11,99]. Furthermore, the exact function of INTS10/13/14 remains undetermined. It was proposed to be involved in DNA/RNA binding and to be required for enhancer binding together with other accessory factors [54]. However, it is not clear how the function of this module corresponds to the activity of the

phosphatase and the nuclease modules and whether this module is required for other Integrator functions.

Finally, Integrator's activity in the context of Herpes viruses seems to be largely independent of the transcription machinery. This *modus operandi* stands in contrast to our current understanding of Integrator activity. It remains unclear how Integrator achieves substrate specificity and nuclease activation in the absence of RNAPII and its pausing factor SPT5 (DSIF). In either case, it is possible that yet undiscovered factors are necessary to regulate Integrator activity under specific circumstances. Identifying different variants of Integrator modules and their impact on different aspects of the Integrator function will be one of the main goals of future research directions.

Acknowledgments

We thank Isaac Fianu and Patrick Cramer for sharing atomic coordinates of the Integrator-PP2A-PEC complex and Alessandro Gardini, Kinga Kamieniarz-Gdula, Tara Best, and Michal Razew for critical comments on the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

W.P.G. was supported by the European Research Council (ERC) grant under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 950278)

ORCID

Moritz M. Pfeleiderer  <http://orcid.org/0000-0002-2369-5824>
Wojciech P. Galej  <http://orcid.org/0000-0001-8859-5229>

References

- [1] Baillat D, Hakimi M-A, Näär AM, *et al.* Integrator, a Multiprotein Mediator of Small Nuclear RNA Processing, Associates with the C-Terminal Repeat of RNA Polymerase II. *Cell*. 2005;123(2):265–276.
- [2] Lai F, Gardini A, Zhang A, *et al.* Integrator mediates the biogenesis of enhancer RNAs. *Nature*. 2015;525(7569):399–403.
- [3] Cazalla D, Xie M, Steitz JA. A Primate Herpesvirus Uses the Integrator Complex to Generate Viral MicroRNAs. *Mol Cell*. 2011;43(6):982–992.
- [4] Rubtsova MP, Vasilkova DP, Moshareva MA, *et al.* Integrator is a key component of human telomerase RNA biogenesis. *Sci Rep*. 2019;9(1). DOI:10.1038/s41598-018-38297-6.
- [5] Xie M, Zhang W, Shu M-D, *et al.* The host Integrator complex acts in transcription-independent maturation of herpesvirus microRNA 3' ends. *Genes Dev*. 2015;29(14):1552–1564.
- [6] Tatomer DC, Elrod ND, Liang D, *et al.* The Integrator complex cleaves nascent mRNAs to attenuate transcription. *Genes Dev*. 2019;33(21–22):1525–1538.
- [7] Elrod ND, Henriques T, Huang K-L, *et al.* The Integrator Complex Attenuates Promoter-Proximal Transcription at Protein-Coding Genes. *Mol Cell*. 2019;76(5):738–752.e7.
- [8] Lykke-Andersen S, Žumer K, Molska EŠ, *et al.* Integrator is a genome-wide attenuator of non-productive transcription. *Mol Cell*. 2021;81(3):514–529.e6.
- [9] Gardini A, Baillat D, Cesaroni M, *et al.* Integrator Regulates Transcriptional Initiation and Pause Release Following Activation. *Mol Cell*. 2014;56(1):128–139.
- [10] Yamamoto J, Hagiwara Y, Chiba K, *et al.* DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nat Commun*. 2014;5(1):4263.
- [11] Fianu I, Chen Y, Dienemann C, *et al.* Structural basis of Integrator-mediated transcription regulation. *Science*. 2021;374(6569):883–887.
- [12] Skaar JR, Ferris AL, Wu X, *et al.* The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Res*. 2015;25(3):288–305.
- [13] Stadelmayer B, Micas G, Gamot A, *et al.* Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat Commun*. 2014;5(1):5531.
- [14] Zheng H, Jin Q, Qi Y, *et al.* Structural basis of INTAC-regulated transcription. *bioRxiv*. 2021. doi:10.1101/2021.11.29.470345
- [15] Chen J, Ezzeddine N, Waltenspiel B, *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*. 2012;18(12):2148–2156.
- [16] Ren W, Chen H, Sun Q, *et al.* Structural basis of SOSS1 complex assembly and recognition of ssDNA. *Cell Rep*. 2014;6(6):982–991.
- [17] Wu Y, Albrecht TR, Baillat D, *et al.* Molecular basis for the interaction between Integrator subunits IntS9 and IntS11 and its functional importance. *Proc Natl Acad Sci USA*. 2017;114(17):4394–4399.
- [18] Pfeleiderer MM, Galej WP. Structure of the catalytic core of the Integrator complex. *Mol Cell*. 2021;81(6):1246–1259.e8.

- [19] Sabath K, Stäubli ML, Marti S, *et al.* INTS10–INTS13–INTS14 form a functional module of Integrator that binds nucleic acids and the cleavage module. *Nat Commun.* **2020**;11(1):3422.
- [20] Zheng H, Qi Y, Hu S, *et al.* Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. *Science.* **2020**;370(6520). DOI:10.1126/science.abb5872.
- [21] Marzluff WF, Koreski KP. Birth and Death of Histone mRNAs. *Trends Genet.* **2017**;33(10):745–759.
- [22] Kumar A, Clerici M, Muckenfuss LM, *et al.* Mechanistic insights into mRNA 3'-end processing. *Curr Opin Struct Biol.* **2019**;59:143–150.
- [23] Sun Y, Hamilton K, Tong L. Recent molecular insights into canonical pre-mRNA 3'-end processing. *Transcription.* **2020**;11(2):83–96.
- [24] Wilusz JE, Spector DL. An unexpected ending: noncanonical 3' end processing mechanisms. *RNA.* **2010**;16(2):259–266.
- [25] Hernandez N. Formation of the 3' end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. *EMBO J.* **1985**;4(7):1827–1837.
- [26] Guiro J, Murphy S. Regulation of expression of human RNA polymerase II-transcribed snRNA genes. *Open Biol.* **2017**;7(6):170073.
- [27] Cazalla D, Yario T, Steitz JA. Down-Regulation of a Host MicroRNA by a Herpesvirus saimiri Noncoding RNA. *Science.* **2010**;328(5985):1563–1566.
- [28] O'Reilly D, Kuznetsova OV, Laitem C, *et al.* Human snRNA genes use polyadenylation factors to promote efficient transcription termination. *Nucleic Acids Res.* **2014**;42(1):264–275.
- [29] Davidson L, Francis L, Eaton JD, *et al.* Integrator-Dependent and Allosteric/Intrinsic Mechanisms Ensure Efficient Termination of snRNA Transcription. *Cell Rep.* **2020**;33(4):108319.
- [30] Nojima T, Tellier M, Foxwell J, *et al.* Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. *Mol Cell.* **2018**;72(6):970–984. e7.
- [31] Berkuyrek AC, Furlan G, Lampersberger L, *et al.* The RNA polymerase II subunit RPB-9 recruits the integrator complex to terminate *Caenorhabditis elegans* piRNA transcription. *EMBO J.* **2021**;40(5):e105565.
- [32] Beltran T, Pahita E, Ghosh S, *et al.* Integrator is recruited to promoter-proximally paused RNA Pol II to generate *Caenorhabditis elegans* piRNA precursors. *EMBO J.* **2021**;40(5). DOI:10.15252/embj.2020105564
- [33] Gómez-Orte E, Sáenz-Narciso B, Zheleva A, *et al.* Disruption of the *Caenorhabditis elegans* Integrator complex triggers a non-conventional transcriptional mechanism beyond snRNA genes. *PLoS Genet.* **2019**;15(2):e1007981.
- [34] Schmidt D, Reuter H, Hüttner K, *et al.* The Integrator complex regulates differential snRNA processing and fate of adult stem cells in the highly regenerative planarian *Schmidtea mediterranea*. *PLoS Genet.* **2018**;14(12):e1007828.
- [35] Corden JL. RNA polymerase II C-terminal domain: tethering transcription to transcript and template. *Chem Rev.* **2013**;113(11):8423–8455.
- [36] Zaborowska J, Egloff S, Murphy S. The pol II CTD: new twists in the tail. *Nat Struct Mol Biol.* **2016**;23:771–777.
- [37] Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet.* **2012**;13(10):720–731.
- [38] Vos SM, Farnung L, Urlaub H, *et al.* Structure of paused transcription complex Pol II-DSIF-NELF. *Nature.* **2018**;560(7720):601–606.
- [39] Zhou Q, Li T, Price DH. RNA Polymerase II Elongation Control. *Annu Rev Biochem.* **2012**;81(1):119–143.
- [40] Egloff S, Szczepaniak SA, Dienstbier M, *et al.* The Integrator Complex Recognizes a New Double Mark on the RNA Polymerase II Carboxyl-terminal Domain. *J Biol Chem.* **2010**;285(27):20564–20569.
- [41] Shah N, Maqbool MA, Yahia Y, *et al.* Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals. *Mol Cell.* **2018**;69(1):48–61.e6.
- [42] West S, Gromak N, Proudfoot NJ. Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature.* **2004**;432(7016):522–525.
- [43] Huang K-L, Jee D, Stein CB, *et al.* Integrator Recruits Protein Phosphatase 2A to Prevent Pause Release and Facilitate Transcription Termination. *Mol Cell.* **2020**;80(2):345–358.e9.
- [44] Cho H, Kim T-K, Mancebo H, *et al.* A protein phosphatase functions to recycle RNA polymerase II. *Genes Dev.* **1999**;13(12):1540–1552.
- [45] Hsin J-P, Manley JL. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* **2012**;26(19):2119–2137.
- [46] Erickson B, Sheridan RM, Cortazar M, *et al.* Dynamic turnover of paused Pol II complexes at human promoters. *Genes Dev.* **2018**;32(17–18):1215–1225.
- [47] Krebs AR, Imanci D, Hoerner L, *et al.* Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Mol Cell.* **2017**;67(3):411–422.e4.
- [48] Steurer B, Janssens RC, Geverts B, *et al.* Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. *Proc Natl Acad Sci U S A.* **2018**;115(19):E4368–E4376.
- [49] Kamieniarz-Gdula K, Proudfoot NJ. Transcriptional Control by Premature Termination: a Forgotten Mechanism. *Trends Genet.* **2019**;35(8):553–564.
- [50] Merino E, Yanofsky C. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends Genet.* **2005**;21(5):260–264.
- [51] Evans R, Weber J, Ziff E, *et al.* Premature termination during adenovirus transcription. *Nature.* **1979**;278(5702):367–370.

- [52] Porrua O, Libri D. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol.* 2015;16(3):190–202.
- [53] Naville M, Gautheret D. Premature terminator analysis sheds light on a hidden world of bacterial transcriptional attenuation. *Genome Biol.* 2010;11(9):R97.
- [54] Barbieri E, Trizzino M, Welsh SA, *et al.* Targeted enhancer activation by a subunit of the Integrator complex. *Mol Cell.* 2018;71(1):103–116.e7.
- [55] Vervoort SJ, Welsh SA, Devlin JR, *et al.* The PP2A-Integrator-CDK9 axis fine-tunes transcription and can be targeted therapeutically in cancer. *Cell.* 2021;184(12):3143–3162.
- [56] Ezzeddine N, Chen J, Waltenspiel B, *et al.* A Subset of Drosophila Integrator Proteins Is Essential for Efficient U7 snRNA and Spliceosomal snRNA 3'-End Formation. *Mol Cell Biol.* 2011;31(2):328–341.
- [57] Beckedorff F, Blumenthal E, daSilva LF, *et al.* The Human Integrator Complex Facilitates Transcriptional Elongation by Endonucleolytic Cleavage of Nascent Transcripts. *Cell Rep.* 2020;32(3):107917.
- [58] Rosa-Mercado NA, Zimmer JT, Apostolidi M, *et al.* Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcriptional despite widespread transcriptional repression. *Mol Cell.* 2021;81(3):502–513.e4.
- [59] Dasilva LF, Blumenthal E, Beckedorff F, *et al.* Integrator enforces the fidelity of transcriptional termination at protein-coding genes. *Sci Adv.* 2021;7(45):eabe3393.
- [60] Wlodarchak N, Guo F, Satyshur KA, *et al.* Structure of the Ca²⁺-dependent PP2A heterotrimer and insights into Cdc6 dephosphorylation. *Cell Res.* 2013;23(7):931–946.
- [61] Virshup DM, Shenolikar S. From Promiscuity to Precision: protein Phosphatases Get a Makeover. *Mol Cell.* 2009;33(5):537–545.
- [62] Albrecht TR, Shevtsov SP, Wu Y, *et al.* Integrator subunit 4 is a 'Symplekin-like' scaffold that associates with INTS9/11 to form the Integrator cleavage module. *Nucleic Acids Res.* 2018;46(8):4241–4255.
- [63] Skaar JR, Richard DJ, Saraf A, *et al.* INTS3 controls the hSSB1-mediated DNA damage response. *J Cell Biol.* 2009;187(1):25–32.
- [64] Li Y, Bolderson E, Kumar R, *et al.* HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J Biol Chem.* 2009;284(35):23525–23531.
- [65] Huang J, Gong Z, Ghosal G, *et al.* SOSS complexes participate in the maintenance of genomic stability. *Mol Cell.* 2009;35(3):384–393.
- [66] Li J, Ma X, Banerjee S, *et al.* Structural basis for multifunctional roles of human Ints3 C-terminal domain. *J Biol Chem.* 2020;296(100112).
- [67] Jia Y, Cheng Z, Bharath SR, *et al.* Crystal structure of the INTS3/INTS6 complex reveals the functional importance of INTS3 dimerization in DSB repair. *Cell Discov.* 2021;7(1):66.
- [68] Lin M-H, Jensen MK, Elrod ND, *et al.* Inositol hexakisphosphate is a critical regulator of Integrator assembly and function. *bioRxiv.* 2021. <http://bioRxiv.org/lookup/doi/10.1101/2021.09.14.460315>
- [69] Dominski Z, Yang X, Purdy M, *et al.* A CPSF-73 Homologue Is Required for Cell Cycle Progression but Not Cell Growth and Interacts with a Protein Having Features of CPSF-100. *Mol Cell Biol.* 2005;25(4):1489–1500.
- [70] Clerici M, Faini M, Aebersold R, *et al.* Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex. *eLife.* 2017;6:e33111.
- [71] Zhang F, Ma T, Yu X. A core hSSB1-INTS complex participates in the DNA damage response. *J Cell Sci.* 2013;126(Pt 21):4850–4855.
- [72] Vos SM, Farnung L, Linden A, *et al.* Structure of complete Pol II-DSIF-PAF-SPT6 transcription complex reveals RTF1 allosteric activation. *Nat Struct Mol Biol.* 2020;27(7):668–677.
- [73] Richard DJ, Bolderson E, Cubeddu L, *et al.* Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. *Nature.* 2008;453(7195):677–681.
- [74] Boeing S, Williamson L, Encheva V, *et al.* Multiomic Analysis of the UV-Induced DNA Damage Response. *Cell Rep.* 2016;15(7):1597–1610.
- [75] Vidhyasagar V, He Y, Guo M, *et al.* Biochemical characterization of INTS3 and C9ORF80, two subunits of hNABP1/2 heterotrimeric complex in nucleic acid binding. *Biochem J.* 2018;475(1):45–60.
- [76] Marzluff WF, Duronio RJ. Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. *Curr Opin Cell Biol.* 2002;14(6):692–699.
- [77] Sun Y, Zhang Y, Aik WS, *et al.* Structure of an active human histone pre-mRNA 3'-end processing machinery. *Science.* 2020;367(6478):700–703.
- [78] Zhang Y, Sun Y, Shi Y, *et al.* Structural Insights into the Human Pre-mRNA 3'-End Processing Machinery. *Mol Cell.* 2020;77(4):800–809.e6.
- [79] Hill CH, Boreikaitė V, Kumar A, *et al.* Activation of the Endonuclease that Defines mRNA 3' Ends Requires Incorporation into an 8-Subunit Core Cleavage and Polyadenylation Factor Complex. *Mol Cell.* 2019;73(6):1217–1231.e11.
- [80] Rodríguez-Molina JB, O'Reilly FJ, Sheekey E, *et al.* Mpe1 senses the polyadenylation signal in pre-mRNA to control cleavage and polyadenylation. *bioRxiv.* 2021. <http://bioRxiv.org/lookup/doi/10.1101/2021.09.02.458805>
- [81] Boreikaite V, Elliott TS, Chin JW, *et al.* RBBP6 activates the pre-mRNA 3' end processing machinery in humans. *Genes Dev.* 2022;36(3–4):210–224.
- [82] Schmidt M, Kluge S, Sandmeier F, *et al.* Reconstitution of ATP-dependent 3'-cleavage of mammalian pre-mRNA reveals a central role of RBBP6. *Genes Dev.* 2022;36(3–4):195–209.

- [83] Xiang K, Nagaike T, Xiang S, *et al.* Crystal structure of the human symplekin–Ssu72–CTD phosphopeptide complex. *Nature*. 2010;467(7316):729–733.
- [84] Mendoza-Figueroa MS, Tatomer DC, Wilusz JE. The Integrator Complex in Transcription and Development. *Trends Biochem Sci*. 2020;45(11):923–934.
- [85] Zhang Y, Koe CT, Tan YS, *et al.* The Integrator Complex Prevents Dedifferentiation of Intermediate Neural Progenitors back into Neural Stem Cells. *Cell Rep*. 2019;27(4):987–996.e3.
- [86] Kapp LD, Abrams EW, Marlow FL, *et al.* The Integrator Complex Subunit 6 (Ints6) Confines the Dorsal Organizer in Vertebrate Embryogenesis. *PLoS Genet*. 2013;9(10):e1003822.
- [87] Otani Y, Nakatsu Y, Sakoda H, *et al.* Integrator complex plays an essential role in adipose differentiation. *Biochem Biophys Res Commun*. 2013;434:197–202.
- [88] Zhang P, Sui P, Chen S, *et al.* INTS11 regulates hematopoiesis by promoting PRC2 function. *Sci Adv*. 2021;7(36):eabh1684.
- [89] Bult CJ, Blake JA, Smith CL, *et al.* Mouse Genome Database (MGD) 2019. *Nucleic Acids Res*. 2019;47(D1):D801–D806.
- [90] Oegema R, Baillat D, Schot R, *et al.* Human mutations in integrator complex subunits link transcriptome integrity to brain development. *PLoS Genet*. 2017;13(5):e1006809.
- [91] Krall M, Htun S, Schnur RE, *et al.* Biallelic sequence variants in INTS1 in patients with developmental delays, cataracts, and craniofacial anomalies. *Eur J Hum Genet*. 2019;27(4):582–593.
- [92] Zhang X, Wang Y, Yang F, *et al.* Biallelic INTS1 Mutations Cause a Rare Neurodevelopmental Disorder in Two Chinese Siblings. *J Mol Neurosci*. 2020;70(1):1–8.
- [93] Federico A, Rienzo M, Abbondanza C, *et al.* Pan-Cancer Mutational and Transcriptional Analysis of the Integrator Complex. *Int J Mol Sci*. 2017;18(5):936.
- [94] Hanahan D, Weinberg RA. Hallmarks of Cancer: the Next Generation. *Cell*. 2011;144(5):646–674.
- [95] Rhodes DR, Chinnaiyan AM. Integrative analysis of the cancer transcriptome. *Nat Genet*. 2005;37(S6):S31–S37.
- [96] Wieland I, Arden KC, Michels D, *et al.* Isolation of DICE1: a gene frequently affected by LOH and down-regulated in lung carcinomas. *Oncogene*. 1999;18(32):4530–4537.
- [97] Filleur S, Hirsch J, Wille A, *et al.* INTS6/DICE1 inhibits growth of human androgen-independent prostate cancer cells by altering the cell cycle profile and Wnt signaling. *Cancer Cell Int*. 2009;9(1):28.
- [98] Li W-J, Hu N, Su H, *et al.* Allelic loss on chromosome 13q14 and mutation in deleted in cancer 1 gene in esophageal squamous cell carcinoma. *Oncogene*. 2003;22(2):314–318.
- [99] Chen J, Waltenspiel B, Warren WD, *et al.* Functional Analysis of the Integrator Subunit 12 Identifies a Microdomain That Mediates Activation of the Drosophila Integrator Complex. *J Biol Chem*. 2013;288(7):4867–4877.