HHS Public Access

Author manuscript

Nat Rev Mol Cell Biol. Author manuscript; available in PMC 2023 September 01.

Published in final edited form as:

Nat Rev Mol Cell Biol. 2023 March ; 24(3): 204–220. doi:10.1038/s41580-022-00534-2.

Genomic Regulation of Transcription and RNA Processing by the Multitasking Integrator Complex

Sarah A. Welsh,

Alessandro Gardini

The Wistar Institute, Philadelphia, PA, USA

Abstract

In higher eukaryotes, fine-tuned activation of protein-coding genes and many non-coding RNAs (ncRNAs) pivots around the regulated activity of RNA polymerase II (Pol II). The Integrator complex is the only Pol II-associated large multi-protein complex that is metazoan-specific and has therefore been understudied for years. Integrator comprises at least 14 subunits, which are grouped into distinct functional modules. The phosphodiesterase activity of the core catalytic module is co-transcriptionally directed against several RNA species, including long non-coding RNAs (lncRNAs), U small nuclear RNAs (U snRNAs), PIWI-interacting RNAs (piRNAs), enhancer RNAs (eRNAs), and nascent pre-mRNAs. Processing of ncRNAs by Integrator is essential for their biogenesis, and at protein-coding genes Integrator is a key modulator of Pol II promoter-proximal pausing and transcript elongation. Recent studies have identified an Integrator-specific protein phosphatase 2A (PP2A) module, which targets Pol II and other components of the basal transcription machinery. In this Review, we discuss how the activity of Integrator regulates transcription, RNA processing, chromatin landscape, and DNA repair. We also discuss the diverse roles of Integrator in development and tumorigenesis.

Introduction

The eukaryotic basal transcription machinery revolves around the enzymatic activity of three distinct RNA polymerases[1]. Shortly after the initial isolation of RNA polymerase II (Pol II) holoenzyme comprising its 12 core subunits[2], it became apparent that efficient transcription initiation and transcript elongation required a host of nuclear accessory factors[3]. In fact, the assembly of the Pol II holoenzyme into a pre-initiation complex (PIC) *in vivo* and *in vitro* requires six multi-subunit general transcription factors (GTFs)[4–6], and the co-activator Mediator complex (26 subunits in mammals)[7–9]. The ensuing dynamics of protein allosteric changes, post-translational modifications and dissociation of PIC components that occur

Corresponding author: A.G. agardini@wistar.org.

Author contributions

Both authors researched data for the article and substantially contributed to discussion of the content. A.G. wrote the article and reviewed and/or edited the manuscript before submission.

Competing interests

The authors declare no competing interests.

Related links

The Cancer Genome Atlas: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga

during the earliest steps of transcription is termed *transcription initiation* and constitutes the main rate-limiting step of protein-coding gene transcription in lower eukaryotes, such as the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*.

Additional mechanisms and protein complexes that regulate gene expression post-initiation have evolved in multicellular organisms. In metazoans, an arrest of Pol II activity occurs at most protein coding genes (and some non-coding RNA (ncRNA) loci) immediately downstream to the transcription initiation site. This so-called 'pausing' of Pol II is a reversible state observed in most metazoans, which deadlocks Pol II while it is engaged in the initial transcription of a nascent, capped pre-mRNA[10–13]. The paused holoenzyme derives from the PIC[14], yet is distinct from the initiating holoenzyme: while parting ways with Mediator and several GTFs, Pol II associates with DRB-sensitivity inducing factor (DSIF) and negative elongation factor (NELF)[15,16]. DSIF and NELF support pausing in multiple ways, for instance by preventing the association of core Pol II subunits with elongation factors such as the PAF complex and SPT6, and blocking the ribonucleotide entry funnel[17,18]. Cyclin-dependent kinase 9 (CDK9), the catalytic subunit of positive transcription elongation factor b (pTEFb), targets several components of the paused Pol II to promote release of the holoenzyme from the transcription start site (TSS) region and elicit productive transcript elongation[15,19,20]. Most notably, CDK9 phosphorylates NELF subunits (thereby promoting its dissociation), DSIF (triggering allosteric changes), and the C-terminal domain **[G]** (CTD) of Pol II at its Ser2 residues (Box 1).

Integrator was serendipitously discovered in 2005 as a new multi-subunit complex in human cells, capable of binding the Pol II CTD (Box 1)[21]. Mass spectrometry revealed peptides from 12 uncharacterized open reading frames associated to several Pol II subunits (Fig. 1). Orthologs for all subunits were identified throughout metazoans, but not in yeast, suggesting that the complex is unique to multicellular eukaryotes. Sequence homology-based annotation tools revealed a MBL/β-CASP **[G]** region within Integrator complex subunit 9 (INTS9) and INTS11, which are highly homologous to cleavage and polyadenylation specificity factor subunit 73 (CPSF73; also known as CPSF3) and CPSF100 (also known as CPSF2), respectively[22], providing the first hint that INTS9 and INTS11 may be endowed with RNA endonuclease activity. In fact, depletion of either the largest subunit (INTS1) or the putative catalytic core (INTS11) of Integrator resulted in specific accumulation of unprocessed, precursor U small nuclear RNAs **[G]** (U snRNAs)[21]. U snRNA processing is known to occur co-transcriptionally, thereby connecting Integrator to Pol II activity. Thus, the complex responsible for cleaving U snRNA transcripts in metazoans was found at last.

The crucial role of Integrator in U snRNA processing is conserved in many organisms, as first shown in 2011 in *Drosophila melanogaster*[23], followed by studies conducted in the nematode *Caenorhabditis elegans*[24] and the highly regenerative planarian *Schmidtea mediterranea*[25]. Orthologs of INTS9 and INTS11 have also been identified in plants[26], but not in fungi, further corroborating the hypothesis that the Integrator complex is the focal point of a transcription regulation network that supports multicellularity. Beyond the original core of 12 subunits, Asunder (INTS13) and the von Willebrand factor A domain containing 9 (INTS14)[27] were identified as putative components of the Integrator complex using gene

reporter assays, and later confirmed to be functionally and biochemically associated to Integrator[28–30]. More recently, serine/threonine-protein phosphatase 2A (PP2A) has been found to stably associate with the Integrator complex, revealing new functions for Integrator in modulating transcription of protein coding genes through dephosphorylation of the Pol II CTD and of Pol II associated complexes such as DSIF[31–34].

From the initial focus on a handful of U snRNA genes, a decade of research has unveiled how Integrator is a genome-wide orchestrator of transcription regulation and RNA processing, being recruited to most protein coding genes and ncRNA loci. In this Review, we discuss the catalytic and non-catalytic functions of the Integrator complex, how they contribute to genome regulation, and how Integrator's malfunction contributes to tumorigenesis in somatic cells and leads to a range of developmental disorders in germinal cells. Biochemical and structural evidence reveal Integrator as a modular, versatile molecular machine that can operate in either Pol II-dependent or Pol II-independent manners. The versatility of Integrator is based on a shared handle and adaptions to different tasks in multiple biological processes.

Structural insights into Integrator

A deeper understanding of the intricate biochemical processes that contribute to gene expression requires structural information on the many involved protein complexes. For instance, cryo-EM analysis of the transcription initiation machinery helped deconvolute how the Mediator complex operates during the earlier steps of transcription[35,36]. Although the architecture of Integrator is far less characterized than that of Mediator, the recent influx of crystallography and cryo-EM data is depicting an outline of this large and elusive complex[28,29,32,33] (Fig. 2).

Endonucleolytic core and backbone components

The sequence similarity between INTS11–INTS9 and CPSF2–CPSF3 led to speculation of close structural analogies between these two endonucleolytic modules recruited by Pol II[21,22,37,38]. In fact, the catalytic core of Integrator adopts a conformation strikingly similar to that of CSPF: the MBL/β-CASP domains at the N-termini of INTS11 and INTS9 form two pseudo-symmetric lobes with outward-facing catalytic pockets (the catalytic site of INTS9 is inactive as it lacks zinc-binding ability)[28,33]. Unlike existing CSPF structures, INTS11 was captured in an apparent inactive conformation, as the cleft leading to the active site is too narrow to accommodate RNA[28,33]. However, there is evidence that binding of Integrator to the paused Pol II elongation complex containing NELF and DSIF, may elicit opening of the catalytic cleft[32].

INTS4, the structural homologue of symplekin in the CPSF complex, folds around the catalytic lobes with its N-terminal HEAT repeats **[G]** and stabilizes their interaction (Fig. 2a). In fact, INTS4, INTS9 and INTS11 form a stable trimeric complex when co-expressed in insect cells[28,39]. The C-termini of INTS11 and INTS9 comprise a robust dimerization domain with multiple contact points. In addition to stabilizing the heterodimer, INTS4 anchors the endonucleolytic module to the rest of the Integrator complex through two

separate domains: its N-terminal HEAT repeats contact the largest subunit, INTS1, while a C-terminal helix bundle makes extensive contact with INTS7 (Ref.[32,33]) (Fig. 1; Fig. 2b).

INTS1, INTS7 and INTS2 form the central backbone of the Integrator complex, making extensive contact with one another. During the biogenesis of Integrator, INTS1, INTS2 and INTS7 may form a nucleating core (Fig. 2a), onto which additional modular components assemble (Fig. 2b). Whereas the INTS2 and INTS7 structures were nearly fully recovered, only about a third of the large 220 kDa INTS1, comprising its C-terminus, was initially solved[33]. Ensuing analyses of Integrator bound to Pol II yielded the structure of an additional central domain of INTS1, which binds INTS2, and captured the mobile N-terminal region of INTS1 through crosslinking mass spectrometry[32]. When bound to the paused Pol II elongation complex, Integrator makes extensive contacts with core Pol II subunits (RPB2 is contacted by an INTS1 N-terminus region, and RPB3 and RPB11 are engaged by INTS7 and INTS2, respectively), with DSIF (INTS11 interacts with SPT5) and with NELF (INTS6 interacts with NELF-B)[32]. Overall, Integrator wraps around the reconstituted paused Pol II through multiple contact points, including at the POL II CTD of RPB1 (Ref.[32]). The Integrator endonuclease module is positioned close to the RNA exit site of Pol II, ready to accommodate and cleave nascent RNA about 20 nucleotides away from the active site[32].

The Int–PP2A module and the INTAC complex

The isolation of the Integrator–PP2A complex (INTAC), from nuclear extracts of human 293T cells, revealed that subunits INTS5 and INTS8 are assembled on top of INTS1, INTS2 and INTS7 to form a large, T-shaped scaffold, thereby anchoring the newly identified phosphatase module of Integrator (Int–PP2A)[33]. INTS5 and, particularly, INTS8 are crucial for tethering components of the phosphatase PP2A to Integrator[32,33] (Fig. 2b). PP2A is one of the most abundant serine/threonine phosphatases in mammalian cells and is generally found in two different compositions. PP2A dimers are assembled from a scaffold subunit (PP2A-A) comprised of a series of HEAT repeats arranged into a horseshoe-like alpha-solenoid shape, and a small globular catalytic subunit (PP2A-C)[40]. Although they are stable *in vitro*, the heterodimers are not considered the functional phosphatase conformation *in vivo*. Instead, PP2A trimers incorporate an additional regulatory subunit (PP2A-B)[41,42], which confers target specificity and ultimately licenses PP2A to dephosphorylate a broad array of protein targets, from mitotic spindle components to intracellular signaling mediators[43].

The structure of INTAC, instead, reveals a PP2A heterodimer tethered to Integrator primarily by INTS8 (and INTS5), which makes extensive contacts with the N-terminal HEAT repeats of PP2A-A[32,33] (Fig. 2b). INTS6 is another central component of Int–PP2A: it binds PP2A (catalytic and scaffold subunits) and helps anchor the phosphatase to the Integrator backbone (Fig. 2b) while simultaneously contacting NELF[32]. Additionally, the INTS6 C-terminal domain may bind INTS3 (Fig. 2a; Fig. 1), although the INTS3 structure has not been solved yet with the fully assembled complex. Deletion of INTS6 results in loss of PP2A components from Integrator[31,33], whereas deletion of INTS8 results in loss of both PP2A components and INTS6 from Integrator[31]. Importantly, integrity of the remaining complex is

preserved following disruption of the Int–PP2A module[31]. Surprisingly, Int–PP2A is devoid of the PP2A-B regulatory subunit, and structural models suggest that INTS6 and INTS8 would clash with its binding[33]. Canonical PP2A complexes are deemed functional only in association with PP2A-B and it is unclear whether Integrator would be sufficient to confer target specificity. Furthermore, the inherent conformational flexibility of PP2A (especially of PP2A-A) and the lack of structural data on most PP2A-B proteins leaves the door open for a role of PP2A-B within Integrator.

The INTS13–INTS14–INTS10 'enhancer' module

INTS13 and INTS14 were not initially recovered as stable Integrator subunits[21]. Subsequent studies identified both proteins as putative Integrator components based on a functional screening in *D. melanogaster*[27]. Human INTS13 associates with the full Integrator–Pol II complex and is also found in a low molecular weight module *in vivo* alongside INTS14 and INTS10 (Ref.[30]) (Fig. 2a). These three subunits can be, in fact, reconstituted *in vitro*[28,29] and, remarkably, INTS13 and INTS14 are found physically interlinked in a heterodimeric conformation that stabilizes both proteins and suggests they are co-translated[29]. INTS10 binds the heterodimer at the N-terminal VWA domain **[G]** of INTS14 (Fig. 1), but its inherent conformational flexibility has precluded further structural analysis. There are structural analogies between a reconstituted INTS13–INTS14 heterodimer and the DNA repair Ku70–Ku80 heterodimer, which binds DNA double-strand break ends[29]. *In vitro* data further suggest that INTS13, INTS14 and INTS10 have some ability to bind single-stranded RNA (ssRNA), double-stranded RNA and ssDNA[29]. Functionally, the INTS13–14–10 module is implicated in transcription regulation as it binds cis-regulatory enhancer elements genome-wide[30], but seems largely dispensable for the endonucleolytic activity of Integrator at U snRNAs[29,30]. Nonetheless, INTS13 appears to be connected to the cleavage module, which it binds through a conserved C-terminal stretch[29].

INTS3 and the SOSS complex

Following INTS1 and INTS2, the third largest subunit of Integrator, INTS3, is a 118kDa protein that is found also in the sensor of ssDNA (SOSS) complex with the small ssDNA-binding proteins nucleic acid-binding protein 1 (NABP1) or NABP2 and with the uncharacterized factor INTS3 and NABP interacting protein (INIP) (Fig. 2a). The SOSS complex is believed to sense ssDNA from DNA damage sites and a partial crystal structure of the complex has been obtained[44]. The N-terminal half of INTS3 acts as a scaffold for the assembly of the ternary complex. Specifically, INTS3 is structured as two alpha helix domains, whose interface generates the C-shaped cavity that docks NABP1 or NABP2. INIP binds to a groove located at the opposite end of INTS3. NABP1 and NABP2, as predicted by the presence of an OB-fold domain **[G]**[45], can associate with short DNA oligos while assembled within the SOSS complex[44]. The SOSS complex also associates with INTS6 or with its paralog INTS6L; these interactions seem to occur through the disordered C-terminal tail of INTS3[46] (Fig. 1; Fig. 2a). The C-terminal half of INTS3, devoid of the disordered tail, was independently expressed in bacteria and crystallized, revealing an elongated HEAT-repeat structure[47,48] (Fig. 1). In addition, there is evidence for homo-dimerization of the INTS3 C-terminal moiety (Fig. 2a). The INTS3 dimer can accommodate the 87 amino acid C-terminal tail of INTS6 (Ref.[47]), which was previously identified as strictly required for

INTS6 binding to SOSS. Additionally, the dimer generates a positively charged groove that has been proposed to bind ssRNA and ssDNA[47], supporting the intrinsic DNA binding ability of NABP1 and NABP2. It remains unclear whether SOSS exists in vivo as an independent complex or is a functional appendix of the fully assembled Integrator complex.

Taken together, the structural data suggest that Integrator is organized in distinct functional modules (Fig. 2). The two known enzymatic activities, endonuclease and phosphatase, latch onto a common scaffold and are physically separated by at least 150 Å. Future studies will elucidate the extent of crosstalk between these two modules. In addition to the catalytic cores, INTS13, INTS14 and INTS10 form a discrete subcomplex (the enhancer module) and INTS3 assembles with two small proteins implicated in DNA repair to form the SOSS complex. INTS6 is an essential component of the Int-PP2A phosphatase module and has also been reported as a SOSS component. Intriguingly, density maps of INTS3 assembled with the full Integrator could not be obtained, suggesting a high degree of INTS3 conformational flexibility. Additional efforts are needed to understand the conformation of this elusive subunit and its precise involvement in other Integrator modules.

Additional experiments are also needed to clarify how the enhancer module is recruited to the larger Integrator complex and its influence on the endonucleolytic module. Lastly, structural information is lacking for INTS12, a low molecular weight Integrator subunit that contains a PHD finger and thus may endow the complex with the ability to read histone methylation patterns.

Processing of non-coding RNAs

RNA cleavage is a key function of the Integrator complex. Like the CPSF complex, Integrator targets nascent RNAs and releases them from the transcription bubble. In this section, we discuss how the activity of Integrator is directed to a broad range of long and short non-coding RNAs[49] (Table 1).

The U snRNA termination machinery

U snRNAs are short (100–200 bp) uridine-rich non-coding RNAs that have essential roles in pre-mRNA splicing. The U snRNAs U1, U2, U4 and U5 are embedded in the spliceosome, whereas U3 and U7 are implicated in rRNA processing and in transcription termination of histone mRNAs, respectively[50–52]. Integrator has emerged as the key U snRNA transcription termination machinery, mirroring the role of the CPSF at protein-coding genes. In fact, the endonuclease module of Integrator (Fig. 2a) has significant homology with CPSF[21]. Elongating Pol II approaching the 3’-end of a gene, encounters an adenine-rich poly(A)-tail signal, which triggers CTD-dependent recruitment of CPSF along with the cleavage stimulation factor (CstF) complex. The CPSF2–CPSF3 heterodimer is endowed with RNA endonucleolytic activity[53]: the zinc-dependent MBL/β-CASP domain of CPSF3 releases the capped nascent mRNA for immediate polyadenylation[37]. The active MBL/β-CASP domain of Integrator within INTS11 shares an overall 40% identity and 60% similarity with that of CPSF3. Initial experiments showed that Integrator is recruited at U1 and U2 loci (both at the promoter and around the cleavage site), and that depletion of INTS11 and INTS1 resulted in accumulation of unprocessed, precursor U1 and U2 transcripts[21]. U snRNA processing

is dependent on the catalytic site of INTS11 (Ref.[21]). Although lacking catalytic activity on its own, INTS9 forms a tight heterodimer with INTS11[28,38] and is essential for proper transcription termination of U snRNAs[38]. Disrupting the interaction surface of the INTS9–INTS11 heterodimer results in defective endonucleolytic activity[38]. The scaffold subunits INTS4[28,39] (Fig. 2a) is also functionally required for U snRNA termination (Fig. 3a). Unlike CPSF[53], the catalytic activity of Integrator has yet to be translated into a standardized biochemical assay, limiting our ability to dissect the contribution of additional subunits to the reaction.

U snRNAs have a peculiar locus architecture, which includes unique DNA motifs that are required for recruiting Pol II: a distal sequence element (DSE) and a proximal sequence element (PSE) relative to the transcription initiation site[54] (Fig. 3a). DSE and PSE are also necessary for transcription termination, suggesting that Integrator must be loaded with Pol II at the initiation site[55,56] (Fig. 3a). Furthermore, an AT-rich sequence of about 14–16 nucleotides, termed the 3' box, is consistently found in human snRNA loci[57]. The 3' box is essential for U1 termination and is positioned immediately downstream of the processed 3' end[52]. Upon depletion of either INTS11 or INTS1, lack of cleavage activity leads to accumulation of unprocessed transcripts spanning the 3' box[21]. The cleavage defects observed following loss of Integrator include Pol II downstream-readthrough and aberrant accumulation of Pol II on chromatin up to few kilobases past the 3' end of U snRNA loci[58]. In organisms with high gene density, such as *C. elegans*, readthrough of U snRNA genes can lead to aberrant transcription of downstream protein-coding genes[59] and can generally be regarded as a signal of environmental stress, leading to heat shock response[24]. Although Integrator is the preeminent machinery of U snRNA transcription termination, CSPF complexes may be able to partially compensate for its absence. Furthermore, stochastic cleavage and termination of U snRNA may occur at some loci in a sequence-dependent manner[60].

Genome-wide data suggest that Integrator is recruited to U snRNA loci co-transcriptionally (at initiation or shortly after transcription initiation)[30,58]. Integrator recruitment occurs, likely, through the Pol II CTD (Box 1) and may be favored by CTD Ser7 phosphorylation[61], although unbiased proteomic studies show that Integrator has comparable affinity for both phosphorylated and unphosphorylated CTD[62]. Overall, it remains unclear how Integrator binds different CTD isoforms — recent data suggest that INTS2, INTS4 and INTS7 may all provide some affinity to the Pol II CTD, especially in the presence of phosphorylated Tyr1 CTD residues[32] (whereas modelling of phosphorylated Ser7 residues shows decreased binding affinity). Other proteins, including DSIF, which is a universal regulator of Pol II pausing at protein-coding genes, may also be important for the recruitment of Integrator to Pol II[63]. Once recruited to transcribed U snRNA loci, it is unclear whether the INTS11 catalytic subunit is present in its active conformation. An intriguing hypothesis is that recognition of the 3' box triggers INTS11 endonucleolytic activity through allosteric mechanisms. Recent reports suggest that the enhancer module INTS13–INTS14–INTS10 may bind to ssRNA containing the U1 3' box[29], which raises the possibility that a dedicated module of Integrator surveys the nascent U snRNA transcript and recognizes the 3' box to further engage the endonucleolytic module. Alternatively, the co-transcriptional formation of the U1 stem-loop structure may elicit binding of INTS13 (Ref.[29]). However, depletion of

INTS13 *in vivo* has little or no effect on U snRNA processing[29,30,64]. Whether Integrator specifically recognizes the 3' box sequence has not been formally demonstrated. Additional factors may also contribute to the 3'-end processing activity of Integrator, including NELF[32,63,65].

U snRNAs are encoded in hundreds of copies scattered across the mammalian genome, though only a subset (~40) are transcribed in any given cell[66]. Nearly all loci are transcribed by Pol II, with the exception of U6 and U6atac, which are Pol III genes[67]. Although ChIP-seq data for Integrator subunits reveal recruitment of the complex across the whole spectrum of Pol II and Pol III U snRNA loci, most studies on Integrator have focused on Pol II. Integrator dynamics and recruitment at U6 sites remains to be elucidated, but it must be noted that Pol II is also found at most Pol III genes[68].

Biogenesis of enhancer RNAs

Developmental processes require temporally-coordinated transcriptome changes that are orchestrated by thousands of enhancer elements[69–72]. Enhancers promote transcription of specific genes by engaging their proximal promoter through the formation of regulatory chromatin loops, which are stabilized by genome architecture modulators such as the cohesin complex and CTCF[73,74]. Enhancer and promoter elements are structurally similar in that their activation requires an accessible, nucleosome-free region, to which sequence-specific transcription factors and the Mediator complex can be recruited[9,69] (Fig. 3b). Similar to promoters, a PIC assembles at enhancers to allow transcription initiation by Pol II[5]. In fact, nearly all active enhancer elements are sites of active Pol II transcription, which is largely bi-directional[75,76] (Fig. 3b). Transcription at enhancer regions produces short-lived non-coding enhancer RNAs (eRNAs), which are relatively short (less than 1kb[75]) and accumulate at chromatin. There are functional implications to eRNAs: they may be required for enhancer-mediated activation of their cognate protein-coding genes[77]. In fact, eRNAs transcription in mammals largely precedes activation of their cognate protein-coding genes[72].

A few eRNAs are polyadenylated by the CPSF complex, which efficiently couples polyadenylation to 3'-end cleavage of the nascent transcript[78,79]. However, the vast majority of eRNAs, which lack a poly(A) tail, is cleaved by the Integrator complex. In fact, Integrator is loaded at active enhancers[30,58,80,81] (Fig. 3b) and its depletion elicits transcriptional readthrough at both 3' ends of bi-directional enhancer loci[58,81,82]. The readthrough is accompanied by accumulation of Pol II downstream of the transcription termination site, similar to what has been observed at U snRNA loci upon Integrator depletion[58]. Furthermore, depletion of Integrator results in increased aberrant polyadenylation of eRNAs, underscoring Integrator being the eRNA termination machinery of choice in physiological conditions[58]. Impaired termination of eRNAs prevents their release from the transcription bubble and may thus impair promoter activation[58,81]. It is unclear whether specific sequences at enhancer loci dictate cleavage. Unlike U snRNAs, enhancer loci are highly diverse in sequence, poorly conserved across species and they lack any motif similar to the 3' box. Nonetheless, the ability of Integrator to generate and maintain a pool of eRNAs

necessary to enforce enhancer–promoter interactions appears fundamental for the execution of stimulus-dependent cellular processes such as differentiation[58,81].

A growing pool of RNA targets

The scope of Integrator's endonucleolytic activity extends beyond U snRNAs and eRNAs (Table 1). In addition to the function of INTS11 at pre-mRNAs and its biological consequences, which are discussed in the next section, many ncRNA species have recently been identified as targets of the Integrator complex in different organisms. Particularly, Integrator has been implicated in post-transcriptional gene silencing mediated by small RNAs such as microRNAs (miRNAs) and PIWI-interacting RNAs (piRNAs). Although miRNAs are best studied in multicellular organisms, they are also encoded by certain viral genomes. For instance, the DNA genome of herpesvirus Saimiri (HVS), a gamma-herpesvirus originally isolated from squirrel monkeys, encodes a handful of protein-coding genes and seven small ncRNAs, whose architecture is similar to human snRNAs[83]. Three out of the seven viral snRNAs contain a stem-loop-forming sequence at the 3' end, well past the 3' box that dictates their termination site. In HVS-infected primate lymphocytes, the host Integrator complex is recruited co-transcriptionally to cleave and release the pre-snRNAs as well as, separately, a downstream stem-loop that functions as an effective primary miRNA (pri-miRNA) transcript[83]. Integrator is then capable of cleaving the 3' end of the pri-miRNA stem-loop, thereby generating a viral precursor-miRNA that is exported to the cytosol to undergo maturation by the host miRNA maturation pathways[84]. Pri-miRNA cleavage by INTS11 appears to be dependent on a distinct, slightly degenerate, A-rich 3' box[84]. Intriguingly, in human cells, depletion of INTS11 (and other Integrator subunits) leads to a global reduction of the pool of endogenous mature miRNAs, but does not affect transcription or processing of the pri-miRNA[85]. Integrator may be required for stabilizing mature miRNA by facilitating their loading onto Argonaute proteins thus possibly favoring their targeting and cleavage of target mRNAs[85].

Similar to miRNAs, piRNAs function in post-transcriptional silencing in association with PIWI (Argonaute family) proteins. Their role is largely restricted to germinal cells, where they surveil transgenerational inheritance by maintaining genome integrity through silencing of transposable elements[86]. In *C. elegans*, piRNAs arise predominantly from two large gene clusters and are transcribed as individual units[86] (Fig. 3c). The 20–40 nucleotides piRNA precursors are transcribed by Pol II; depletion of Integrator's catalytic subunit causes a marked reduction of piRNA abundance and an increase in precursor size[87]. The abundance of piRNA is controlled directly by the catalytic activity of Integrator, rather than by modulation of Pol II elongation[87] (Fig. 3c). Whereas Integrator is generally recruited co-transcriptionally by the Pol II CTD, in nematodes the RPB9 subunit of Pol II may have an additional role in promoting localization of Integrator at piRNA loci *in vivo*[88] (Fig. 3c). Genetic screens implicate Integrator in piRNA biogenesis of *D. melanogaster*[89,90], suggesting that transgenerational silencing of transposable elements may rely on Integrator in all metazoans.

Whereas eRNAs, microRNAs and piRNAs comprise large and diverse classes of transcripts, there are single-copy ncRNAs that function as highly specialized regulators of biological

processes. For instance, maintenance of chromosome ends hinges on the activity of telomerase, a reverse transcriptase capable of synthesizing new DNA ends using as template the long non-coding RNA (lncRNA) telomerase RNA template component (TERC)[91]. Transcription termination of TERC by 3'-end cleavage relies on Integrator and is specified by TERC's promoter, similar to U snRNAs[92]. The ~20kb lncRNA nuclear paraspeckle assembly transcript 1 (NEAT1) is an essential component of paraspeckles, which are stress-related nuclear bodies[93]. In physiological conditions, the endonucleolytic module of Integrator promotes early termination of NEAT1, thereby generating a ~4kb isoform that does not efficiently nucleates paraspeckles[94]. Interestingly, NEAT1 cleavage by Integrator may be coordinated with, rather than antithetical to, recruitment of the CPSF complex, since the short NEAT1 isoform is known to be polyadenylated[95]. Further crosstalk between Integrator and CPSF may occur at the 3'end of select protein-coding genes, as we discuss in the ensuing section. Lastly, the role of Integrator in overseeing transcription termination of ncRNAs[49] might be more extensive that anticipated. In fact, promoter upstream transcripts (PROMPTs) that originate antisense to a large set of protein coding genes appear to be terminated primarily by Integrator[82]. In this regard, the transcription elongation factor SPT6 seems critical to ensure proper recruitment of Integrator at many ncRNA loci, including eRNAs[82].

In summary, an increasing body of work demonstrates that Integrator operates as a 3'-end RNA processing factor for a broad range of long and short ncRNAs in several species (Table 1). The endonucleolytic activity of Integrator is co-transcriptional and maintains homeostatic levels of U snRNAs, TERC, piRNAs, and certain promoter antisense transcripts.

Transcription of protein-coding genes

RNA cleavage by Integrator is not restricted to ncRNAs. In fact, nascent mRNAs are targeted by Integrator, primarily during Pol II pausing and early elongation. Cleavage may also occur during late elongation and around the termination site. Thus, the scope of Integrator's activity at coding genes encompasses the entire transcription process and, in addition to the cleavage module, implicates the newly discovered phosphatase module as a tuner of Pol II processivity.

Co-transcriptional targeting of nascent mRNAs by the endonucleolytic module

The tight control of Pol II activity at protein-coding genes is crucial for executing coordinated gene expression programs that vary across cells, tissues and developmental stages. Following transcription initiation, and before the onset of productive transcript elongation, Pol II undergoes a regulatory step of pausing, which occurs after polymerization of a 30–50nt nascent, capped mRNA[10,11,96–101]. This promoter-proximal Pol II pausing is one of the most conserved mechanisms of gene regulation and co-evolved in multicellular organisms as a rheostat for transcriptional gene regulation during development[102,103]. Nonetheless, the role of Pol II pausing is not limited to development, as adult tissues rely on promoter-proximal Pol II pausing to be able to induce a rapid and coordinated transcriptional response to a variety of extracellular cues[104–108].

Several reports have implicated Integrator in the regulation of Pol II pausing and elongation, with the proposed mechanism varying according to the model system and the Integrator subunits examined. The first genome-wide analysis of Integrator chromatin occupancy revealed widespread association of INTS11 and Pol II at active genes[108]. In human cells, Integrator is highly enriched at Pol II initiation and pausing sites and remains co-transcriptionally associated with Pol II throughout the gene body and the 3'-end of highly active genes[30,31,108,109]. Similarly, INTS3 broadly associates with promoter-proximal Pol II pausing sites, in correlation with NELF and DSIF occupancy[110]. Additional data from *D. melanogaster* cell lines suggest that Integrator may be most enriched at genes with 'unstable' promoter-proximal Pol II pausing, which are marked by fast turnover of paused Pol II and characterized by lower levels of histone H3 Lys4 trimethylation compared to genes with more stable pausing[111].

Integrator appears to have distinct functions in Pol II regulation at different sets of protein-coding genes. For instance, the depletion of INTS11 or INTS1 blunts stimulus-dependent activation of immediate early genes (IEGs)[108]. IEGs accumulate promoter-proximal Pol II and are primarily regulated by a pause-release mechanism. Integrator depletion prevents Pol II release at IEGs upon stimulation[108,109], in a catalytic-dependent manner[109]. An Integrator-dependent pause-release mechanism extends well beyond stimulus-dependent transcription and is shared by a large fraction of genes, while transcribed at steady-state levels[109]. In fact, depletion of Integrator in HeLa cells hampers productive elongation and correlates with Pol II stalling upon reaching the +1 nucleosome, where a second, NELF-independent barrier to elongation is mounted[112]. Notch-dependent transcription in tumor cells also appears to be relying on INTS11, the catalytic subunit of Integrator[113].

By contrast, in lower eukaryotes, Integrator seems particularly prone to suppression of transcription. In fact, depletion of the catalytic core in *D. melanogaster* leads to de-repression of a set of protein-coding genes, whereas a far smaller number of genes are downregulated[112]. In addition to repressing steady-state transcription, stimulus-dependent transcription is also curtailed by Integrator at *D. Melanogaster* copper-induced promoters (where new Pol II initiation occurs)[64], whereas the activation of heat shock genes, which is dependent on pause-release, is enforced by Integrator similar to human IEGs[108].

There are multiple lines of evidence that Integrator directs its catalytic activity against nascent transcripts early during their elongation[64,109,111], albeit with alternative outcomes (Fig. 4). According to one model of transcription activation, mRNA cleavage by the Integrator after 21-nucleotides is conducive to release of paused Pol II, followed by loading of an elongation-competent Pol II complex[109] (Fig. 4a). According to an alternative model of elongation repression, Integrator is preferentially recruited to a subset of genes under promoter-proximal Pol II pausing, where its endonucleolytic activity results in premature transcription termination **[G]** (PTT), therefore preventing paused Pol II complexes from becoming fully licensed for elongation[111] (Fig. 4b). A similar role of 'premature termination' has been described for the CSPF factor PCF11 (Ref.[114]), and PTT has long been considered an important mechanism of transcription regulation[115].

Beyond the Pol II pausing region, the endonucleolytic activity of Integrator may be further used at protein-coding genes for broad-range termination of non-productive or stalled transcription, as far as few kilobases from the TSS[116]. This activity occurs especially at lowly expressed loci, in a way similar to how Integrator restricts the accumulation of certain lncRNAs[82]. Thus, in mammalian cells, Integrator may be a tool to tame pervasive transcription. According to multiple ChIP-seq datasets obtained from human cells, Integrator is tied to transcribing Pol II from initiation through termination. Integrator has been proposed to regulate transcription termination and Pol II positioning at protein-coding genes that have a motif similar to the 3' box of U snRNAs[110] and to mediate termination of mRNAs of replication-dependent histones[117]. Recent data suggest that, although most protein-coding genes are exclusively dependent on the CPA machinery for transcription termination and polyadenylation, a subset of genes enriched in proximal alternative polyadenylation sites require the activity of INTS11 to ensure selection of the proximal site[118]. More generally, the activity of Integrator at transcription termination sites is crucial during cellular stress conditions (i.e. hyperosmotic shock), when transcription downstream of canonical polyadenylation sites become pervasive and is partly attenuated by Integrator activity[119].

The phosphatase module of Integrator

Over the past two years, multiple reports have uncovered a novel mechanism through which Integrator regulates Pol II transcription in association with PP2A. PP2A is a highly conserved, ubiquitously expressed Ser/Thr phosphatase[40,43], which accounts for the majority of phosphatase activity in any given cell[40,120]. *In vivo*, PP2A works primarily as a trimeric complex of a scaffold (PP2A-A), a catalytic (PP2A-C) and a regulatory (PP2A-B) subunit[40]. Regulatory subunits are encoded by at least 15 different genes in human and confer target specificity to the A/C catalytic core[121] (Fig. 5a). The first evidence for an association between the Integrator and PP2A came from comprehensive proteomics studies suggesting the existence of robust protein–protein interactions between submits of the two complexes[122,123]. A decade later, a landmark paper reported the cryo-EM structure of nine subunits of the Integrator complex bound to a PP2A holoenzyme: the catalytic subunit PP2A-C and the scaffold subunit PP2A-A (together termed INTAC)[33]. While striving to obtain a first comprehensive structure of Integrator through overexpression of several tagged subunits followed by affinity purification, the authors also purified stoichiometric amounts of endogenous PP2A, suggesting that most endogenous Integrator may in fact carry the phosphatase module (Fig. 2; Fig. 5a) and the resulting INTAC could be the prevailing Integrator-complex variant in mammalian cells. Interestingly, no PP2A-B subunits were reported to be stably associated with Integrator, and structural data from PP2A heterotrimeric complexes suggests a steric clash between PP2A-B and INTS6 or INTS8 (Ref.[33]). Two additional studies functionally characterized the association of Integrator and PP2A and its biological significance, broadening the scope of action of Integrator in the Pol II transcription process[31,34].

As discussed above, transcription by Pol II is regulated by dynamic phosphorylation of the highly conserved Pol II CTD (Box 1) and of other Pol II-associated factors such as NELF and DSIF[124]. Whereas the kinases involved in this process have been

extensively studied[125,126], an understanding of the relevant phosphatases, especially in higher eukaryotes, has lagged behind largely due to the increased complexity and apparent promiscuity of phosphatase complexes in higher eukaryotes[120]. Progression through the distinct stages of Pol II transcription requires the activity of specific CDKs such as CDK7, CDK9, CDK12 and CDK13 (Ref.[127]). In particular, paused Pol II is targeted by CDK9–cyclin T1 (the p-TEFb complex), which is recruited through the super elongation complex[128] to phosphorylate the Pol II CTD, DSIF, NELF and other Pol II associated factors (i.e. PAF1 and SPT6) at the transcription bubble (Fig. 5b). Recruitment of pTEFb is paramount to release of paused Pol II and transition into a productive elongation phase. Consistent with recruitment of Integrator to active genes, PP2A is also found near the pausing site and at the gene bodies of most (if not all) expressed genes, in multiple cell types[31,33]. Depletion of INTS8 or INTS6 prevents efficient PP2A recruitment to chromatin[31,33] and is broadly associated with increased transcriptional output measured as both nascent RNA and steady state RNA levels[31,33,34]. Phosphorylated DSIF (SPT4–SPT5) is conducive to productive elongation[18], and loss of Int–PP2A leads to increased phosphorylation of SPT5 at Ser666 (Ref.[34,129]) and at Thr806 (Ref.[31,129]), suggesting that Pol II pausing downstream the TSS is the result of a balancing activity between phosphatases and kinases (Fig. 5b).

More specifically, Int–PP2A appears to functionally oppose pTEFb, and INTS6 depletion releases pausing induced by highly specific CDK9 inhibitors (CDK9i)[31]. Furthermore, CDK9i and allosteric activation of PP2A synergize to secure Pol II in a paused deadlock, effectively killing elongation-addicted acute myeloid leukemia cells[31]. Even in the absence of functionality of the elongation factor SPT5, loss of Int–PP2A promotes elongation[129], underscoring its all-around role in enforcing Pol II pausing. In fact, phosphorylated Ser2 CTD residues appear to be another key substrate of Int–PP2A[31] (Fig. 5b). Phosphorylated Ser5 and phosphorylated Ser7 residues may also be targeted by Integrator[31,33,34,129] (Box 1), and ChIP-seq data of PP2A-A and PP2A-C indicate they diffusely interact with chromatin across the body and 3’-end of most genes[31], suggesting that Int–PP2A modulates Pol II activity even beyond pausing. PP2A is not active at Tyr residues and, intriguingly, phosphorylated Tyr1 CTD may be necessary for Integrator binding[32]. In addition to PP2A, PP1–PNUTS and PP4 have been also linked to Pol II regulation, opposing CDK9 (p-TEFb) activity at specific residues throughout the transcription process[130,131]. Remarkably, Integrator and PP1 were found to be biochemically associated[31,122] (Fig. 5b), albeit at a lower stoichiometry compared with PP2A, and the combined PP1 and PP2A inhibition improves rescue of CDK9i-induced pausing over PP2A inhibition alone[31]. The activity of PP1 and PP2A may be interdependent and largely coordinated; the interplay with other phosphatases operating on Pol II (SSU72, RPAP2, FCP1) will need to be further explored[120].

Non-catalytic roles of Integrator

Beyond the endonucleolytic core and the Int–PP2A phosphatase module, Integrator subunits have been shown to assemble into two additional modules with distinct functions: the enhancer module and the SOSS complex.

The INTS13, INTS14 and INTS10 subunits of Integrator can assemble into the smaller, biochemically independent[28–30] enhancer module (Fig 2a). This module was shown to form in vivo in hematopoietic progenitor cells and to bind developmental enhancer elements independently of the main Integrator complex and in the absence of Pol II[30]. None of these three stably associated subunits contains predicted DNA-binding motifs or chromatin-binding domains (Fig. 1). Intriguingly, INTS13 and INTS14 are capable of binding nucleic acids to some extent, thereby providing a putative mechanism for recruitment that is independent of active transcription[29]. Additionally, the transcription factor early growth response protein 1 (EGR1) and its co-factor NAB2 interact with the enhancer module and are required for its recruitment to poised enhancers[30]. Enhancer activation by EGR1–NAB2–INTS13 axis is necessary for monocyte commitment of myeloid progenitor cells, highlighting a new Integrator function at enhancers beyond transcription termination. The enhancer module may effect transcription of developmental genes beyond myeloid cells, as suggested by a report of INTS13 mutations in a subtype of an oral-facial-digital developmental syndrome[132]. Additional work will need to clarify how the enhancer module interlaces with the endonucleolytic module and the rest of Integrator[29,132].

The SOSS complex comprises INTS3 (SOSS-A), NABP1 (SOSS-B1) or NABP2 (SOSS-B2) and the uncharacterized protein INIP (SOSS-C)[133–135]. The three proteins form a stable complex *in vivo* and *in vitro*, in which INTS3 acts as a scaffold[47,48]; the SOSS complex may also include INTS6 or INTS6L[46,48] (Fig. 2a). NABP1 and NABP2 were discovered and studied for their ability to bind ssDNA through their OB fold, similar to the RPA complex that is a well-studied player in DNA replication and DNA repair[136,137]. Generation of ssDNA ends at sites of double–strand DNA break (DSBs) is paramount for DNA repair and genome integrity. During homologous recombination (HR), 3’ overhangs are generated at DSBs and promptly recognized by ssDNA binding proteins, which protect them from exonucleases and initiate the HR process. INTS3 is necessary for recruitment of NABP1 and NABP2 to DSBs[133–135] and supports efficient HR-mediated repair[134,135]. Correspondingly, depletion of any SOSS component sensitizes cells to ionizing radiation (which produces DNA breaks)[44,134,135]. DSBs repaired by HR are first recognized by the MRE11–RAD50–NBS1 (MRN) complex, which recruits and activates the kinase ATM to ultimately coordinate checkpoint activation and recruitment of additional repair proteins. SOSS components are required for recruitment of MRN and for ATM activation[138,139]. INTS3, NABP1 and NABP2, and INIP frequently co-precipitate with various Integrator subunits (Fig. 1); furthermore, INTS6 can associate and work with SOSS and is also assembled in the phosphatase module of Integrator. At this stage, there is no conclusive evidence whether SOSS is an integral part of the main Integrator complex and additional work will have to elucidate the dual role of INTS6 as a recruiter and stabilizer of Int–PP2A and as a recruiter of SOSS to sites of DNA damage.

Integrator in development and disease

As a complex that ubiquitously regulates transcription and RNA processing in metazoans, Integrator is deemed essential across all tissue types and developmental stages. Integrator’s composite architectural and functional modularity originates a nuclear regulatory hub utilized by Pol II, transcription factors and DNA damage response proteins. The pathological

implications of Integrator's malfunction vary according to the subunit and module disrupted, from intellectual disability[140] and leukemogenesis[141] to chronic obstructive pulmonary disease[142]. In this section we discuss the role of Integrator in a subset of inheritable neurodevelopmental disorders, and in multiple tumor types.

Roles of Integrator in development and developmental disorders

Integrator function is required during early development and tissue morphogenesis, as well as cell differentiation in the adult organism. In developing mouse embryos, deletion of *Ints1* results in early lethality, likely by destabilization of the entire complex[143]. Similarly, germ-line mutations of *D. melanogaster* Integrator core components result in mid-to-late larval lethality[23]. In zebrafish embryos, mutation of Ints6 disrupts progression of gastrulation owing to de-repression of dorsal organizer genes, resulting in severe dorsalization[144]. This phenotype is consistent with INTS6-mediated enforcing of Pol II pausing[31]. Integrator is also crucially required in early embryonic stages in *Artemia sinica* (brine shrimp)[145]. Furthermore, a fully functional Integrator complex is deemed essential to coordinate transcriptional programs in later stages of development, as well as during differentiation of adult stem cells. For example, Integrator malfunction compromises hematopoiesis at multiple levels: by downregulating SMAD–BMP signaling[146], failing to activate EGR1-dependent developmental enhancers[30], and de-repressing Polycomb target genes[147]. In the stem cell-rich planarian flatworm, loss of core Integrator subunits disrupts stem-cell maintenance and tissue regeneration[25]. In a mouse model of adipogenesis, differentiation depends on increased expression of Integrator subunits[148]. The role of Integrator in regulating developmental gene expression programs has been further investigated in neuronal development. In mouse neuronal progenitor cells, Integrator coordinates the expression of a set of genes responsible for proper migration of newly formed neurons across the developing cerebral cortex, in coordination with the zinc finger protein Zfp609 and the cohesin loading factor Nibpl[80]. During *D. melanogaster* development, Integrator prevents dedifferentiation of intermediate neural progenitor cells, thereby facilitating terminal cell fate commitment[149]. All the above mechanisms may contribute to a handful of developmental cognitive syndromes that have been recently described (Table 2). Recessive mutations in *INTS8* cause a severe neurodevelopmental disorder characterized by brain malformations, facial dysmorphism and severe cognitive delay[140]. Recessive mutations in *INTS1* have a similar clinical phenotype[140,150,151], with additional characteristic traits such as juvenile cataract (Table 2). Interestingly, all affected individuals show severely impaired or absent speech, perhaps suggesting the inability to sufficiently establish the neural circuits governed by the FOXP1–FOXP2 transcription factors axis[152].

The phenotypes of *INTS8* mutations may shed a light on the role of the Int–PP2A module and Pol II pausing in human development. Interestingly, point mutations in PP2A-A are associated with a wide spectrum of neurodevelopmental disorders[153–157]. We speculate that transcription impairment in developing embryos carrying these PP2A-A mutations may be a major pathogenic driver. Future work on these developmental disorders will have to assess the precise contributions of disrupting canonical PP2A ternary complexes versus Integrator-bound PP2A. In line with the role of INTS1 and INTS8 in human development, indirect disruption of Integrator's activity may underlie another neurodevelopmental syndrome,

termed Galloway–Mowat[158], which is genetically determined by loss-of-function mutations in WDR73 (reported as a novel interactor of Integrator's endonucleolytic module). Lastly, a component of the enhancer module of Integrator, INTS13, has been recently associated with a developmental ciliopathy, resulting in several oro-facial and digital anomalies[132] (Table 2). Primary cilia are microtule-based sensory organelles that protrude from the surface of most cells and transduce and regulate several signaling pathways during development[159]. Integrator may affect centriole localization and ciliogenesis by ensuring correct processing of ncRNAs crucial to the process[160,161] or by maintaining the proper dosage of ciliary gene transcripts through regulation of their promoter-proximal Pol II pausing[132].

Roles of Integrator in tumorigenesis

Ubiquitously-expressed transcription co-activator complexes such as the nucleosome remodeler SWI/SNF have recently emerged as hotbeds of tumor-driving mutations[162]. By comparison, Integrator is less frequently mutated in human cancers and its role in tumorigenesis was not immediately recognized. Analysis of all 15 Integrator subunits (including INTS6L) across a comprehensive panel of primary tumor samples from the Cancer Genome Atlas collection revealed that nonsynonymous mutations in Integrator subunits occur in up to 10% of patients[163]. The highest rate of mutations in all tumor types combined is found in *INTS1*, *INTS2* and *INTS8*, whereas *INTS3* and *INTS7* are frequently mutated in diffuse large B-cell lymphomas and pancreatic adenocarcinomas, respectively[163]. Intriguingly, a pan-tumor analysis highlighted *INTS10* as the gene most subjected to purifying selection during tumor evolution[164]. In essence, tumors bearing one copy of wild-type *INTS10* (hemizygous), preserve this wild type allele by selecting against cells that accumulate missense mutations, underscoring the importance of *INTS10* for tumor progression and suggesting Integrator is a tumor vulnerability[164]. Similarly, the Integrator catalytic core INTS9–INTS11 is rarely mutated in tumors and there are different nuances to its potential role as a disease driver, depending on the tumor stage and origin. On one side, the catalytic module of Integrator promotes expression of a large fraction of genes in mammalian cells, including cell cycle activators and anti-apoptotic proteins[108–110]. Particularly, Integrator acts downstream of MAP kinase signaling by enabling transcriptional responsiveness to MAPK-activating mutations that are frequently found in different tumor types such as melanomas[165]. Inhibition of INTS11 could contribute to curbing tumor cell growth especially in this subset of tumors[165]. On the other side, tumor initiation may benefit from reduced INTS11 activity. The recently identified role of INTS11 in preventing paraspeckles formation thorough suppression of the full-length NEAT1 isoform, suggests that the pro-oncogenic role of these nuclear bodies may also depend on hampered Integrator activity[94]. This role could explain how reduced expression of Integrator components correlates with poor response to chemotherapy and with dismal survival rates in ovarian cancerpatients[94].

Several Integrator subunits are overexpressed in esophageal adenocarcinomas compared to normal epithelial cells; in these tumors, Integrator supports aberrant activation of Notch signalling and is essential for cell growth and tumorigenesis[113]. *INTS7*, *INTS8* and *INTS13* are frequently found upregulated in RNA-seq datasets of primary tumors matched to healthy tissue samples[163]. Upregulation of INTS8 is also specifically associated with epithelial-to-

mesenchymal transition and increased metastatic potential in hepatocellular carcinoma[166]. Interestingly, these subunits carry very different roles within the complex. Expression of another subunit, INTS3, is amplified in some hepatocellular carcinomas[167], but repressed in a subset of acute myeloid leukemias through an aberrant mechanism of intron inclusion, which is brought about by misregulation of serine/arginine-rich splicing factor 2 (SRSF2)-dependent splicing[141] (Table 2). Loss of INTS3 in hematopoietic stem and progenitor cells is crucial for boosting clonal expansion and proliferation, a hallmark of leukemogenesis[141].

Discovery of the Integrator phosphatase module[31] has major implications on tumorigenesis and suggests that part of Integrator may operate universally as a tumor suppressor. Int–PP2A restrains growth of *MLL*-rearranged leukemias and of solid tumors by enforcing Pol II pausing and opposing CDK9-dependent transcript elongation[31]. INTS6 is a key component of Int–PP2A and was originally named deleted in cancer 1 (DICE1), after its propensity to undergo deletion in cancer cells[168]. In fact, loss or downregulation of INTS6 correlates with poor survival and is common in many tumor types[31], including prostate cancer, hepatocellular carcinoma and hematopoietic malignancies[169–171]. PP2A has long been regarded to as a tumor suppressor: activity of its various complexes is reduced in multiple tumor types by way of epigenetic silencing, somatic mutations and upregulation of endogenous inhibitors[172–178]. Small molecule activators of PP2A have been recently developed and have demonstrated broad antitumoral activity[179–181]. Certain class of PP2A activators are effective at increasing Int–PP2A recruitment to chromatin and act synergistically with CDK9 inhibitors to deadlocks a paused Pol II conformation, resulting in reduced tumor burden and increased survival in different mouse models[31].

Conclusions and future directions

Fifteen years after its discovery, Integrator has now come into prominence as a keystone of transcription and co-transcriptional RNA processing in all higher eukaryotes. Universally recruited to Pol II-transcribed loci, likely at the transcription initiation–pausing transition, Integrator surveils the early as well as the late steps of the transcription process. Beyond controlling expression of Pol II genes, Integrator's tasks in genome regulation further span from enhancer activation to sensing DNA damage sites. The biochemical, structural and genetic dissection of the Integrator complex is advancing our overall understanding of biological processes such as ncRNA biogenesis and Pol II pausing. Integrator studies have revealed how RNA cleavage solves polymerase pausing, for better or worse. Additional mechanistic insights will further clarify whether gene activation and repression mediated by INTS11 are mutually exclusive, and whether evolutionary pressure has reshaped the global roles of Integrator. Future studies should ideally develop and use physiologically more relevant experimental systems to model Integrator activity during promoter-proximal Pol II pausing, such as primary human and mouse embryonic-like cells and differentiated cells.

The recent, exciting discovery of Int–PP2A revealed that Integrator is endowed with a second catalytic activity, in addition to being an RNA phosphodiesterase (similar to the CPA complex). Notably, Int–PP2A is the first non-canonical PP2A complex described so far, raising intriguing questions on how canonical PP2A modulators (such as PP2A-B

subunits, endogenous inhibitors, paralogous subunits) intertwine with Integrator and the transcription process. The RNA cleavage and the dephosphorylation activities of PP2A are likely coordinated during transcription, characterization and modelling of which will require extensive structural biology and biochemical investigations. Recent breakthroughs in structural biology have unveiled how Integrator is folded around a paused Pol II holoenzyme, also revealing structural clash points with elongation factors such as PAF and SPT6. Functional and biochemical evidence, however, suggest that Integrator is intimately associated to Pol II well beyond the promoter-proximal pausing site. Future studies of this inherently flexible complex will have to determine how Integrator adapts its conformation to escort polymerase through elongation and termination. Furthermore, the full scope of Integrator's phosphatase activity may extend beyond Pol II (and associated factors) and target transcription factors, chromatin modifiers or even nucleosomes.

Disruptions of Integrator's activity through copy number alterations or point mutations of its subunits are being systematically reported. Integrator is also emerging as a nexus of transcriptional responses to various mitogenic signaling pathways and cellular stress. In this capacity, Integrator may have an essential role in tumor development. Given the broad sphere of action of Integrator at chromatin, there is need for rigorous pathophysiological studies that will pinpoint the specific activity and module of the complex that drive tumorigenesis and developmental disorders. By clarifying the molecular basis of Integrator malfunction, we will better appreciate its potential as a pharmacological target.

Acknowledgements

Work in the Gardini lab is supported by grants from the NIH (R01 HL141326 and CA252223), the American Cancer Society (RSG-18-157-01-DMC) and the G. Harold and Leila Y. Mathers Charitable Foundation (A.G.).

Glossary

U small nuclear RNA
Uridine-rich small nuclear RNAs transcribed by Pol II that have essential roles in pre-mRNA splicing

MBL/β-CASP domain
The metallo-β-lactamase (MBL) fold is shared by a diverse set of enzymes, including a large family of bacterially-derived antibiotic hydrolases. In a subset of beta-lactamase enzymes operating on nucleic acid substrates, the MBL fold further extends into a beta-CASP globular domain to form an active nuclease site

C-terminal domain
The unstructured and highly repetitive C-terminal domain of the largest subunit of Pol II is intricately phosphorylated to regulate Pol II function

HEAT repeats
A protein structural motif composed of tandem repeats of 2 α-helices linked by a short loop

OB-fold domain

The oligonucleotide or oligosaccharide binding fold is an evolutionarily ancient protein domain capable of binding nucleic acids

VWA domain

The von Willebrand factor type A domain is an alternating sequence of α-helices and β-strands, generally mediating protein–protein interactions

Premature transcription termination (PTT)

generally occurs when elongating Pol II is arrested at any point after the TSS and before reaching a canonical termination site, usually resulting in release of unstable transcripts

REFERENCES

1. Roeder RG & Rutter WJ Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature 224, 234–237, doi:10.1038/224234a0 (1969). [PubMed: 5344598]
2. Young RA RNA polymerase II. Annu Rev Biochem 60, 689–715, doi:10.1146/annurev.bi.60.070191.003353 (1991). [PubMed: 1883205]
3. Weil PA, Luse DS, Segall J & Roeder RG Selective and accurate initiation of transcription at the Ad2 major late promotor in a soluble system dependent on purified RNA polymerase II and DNA. Cell 18, 469–484, doi:10.1016/0092-8674(79)90065-5 (1979). [PubMed: 498279]
4. Orphanides G, Lagrange T & Reinberg D The general transcription factors of RNA polymerase II. Genes Dev 10, 2657–2683, doi:10.1101/gad.10.21.2657 (1996). [PubMed: 8946909]
5. Koch F et al. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nat Struct Mol Biol 18, 956–963, doi:10.1038/nsmb.2085 nsmb.2085 [pii] (2011). [PubMed: 21765417]
6. Plaschka C et al. Architecture of the RNA polymerase II-Mediator core initiation complex. Nature 518, 376–380, doi:10.1038/nature14229 (2015). [PubMed: 25652824]
7. Schier AC & Taatjes DJ Structure and mechanism of the RNA polymerase II transcription machinery. Genes Dev 34, 465–488, doi:10.1101/gad.335679.119 (2020). [PubMed: 32238450]
8. Flanagan PM, Kelleher RJ 3rd, Sayre MH, Tschochner H & Kornberg RD A mediator required for activation of RNA polymerase II transcription in vitro. Nature 350, 436–438, doi:10.1038/350436a0 (1991). [PubMed: 2011193]
9. Richter WF, Nayak S, Iwasa J & Taatjes DJ The Mediator complex as a master regulator of transcription by RNA polymerase II. Nat Rev Mol Cell Biol, doi:10.1038/s41580-022-00498-3 (2022).
10. Rahl PB et al. c-Myc regulates transcriptional pause release. Cell 141, 432–445, doi:10.1016/j.cell.2010.03.030 (2010). [PubMed: 20434984]
11. Core LJ, Waterfall JJ & Lis JT Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322, 1845–1848, doi:10.1126/science.1162228 1162228 [pii] (2008). [PubMed: 19056941]
12. Zeitlinger J et al. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet 39, 1512–1516, doi:10.1038/ng.2007.26 (2007). [PubMed: 17994019]
13. Muse GW et al. RNA polymerase is poised for activation across the genome. Nat Genet 39, 1507–1511, doi:10.1038/ng.2007.21 (2007). [PubMed: 17994021]
14. Fant CB et al. TFIID Enables RNA Polymerase II Promoter-Proximal Pausing. Mol Cell 78, 785–793 e788, doi:10.1016/j.molcel.2020.03.008 (2020). [PubMed: 32229306]
15. Kwak H & Lis JT Control of transcriptional elongation. Annu Rev Genet 47, 483–508, doi:10.1146/annurev-genet-110711-155440 (2013). [PubMed: 24050178]
16. Li J et al. Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. Mol Cell 50, 711–722, doi:10.1016/j.molcel.2013.05.016 (2013). [PubMed: 23746353]

17. Vos SM, Farnung L, Urlaub H & Cramer P Structure of paused transcription complex Pol II-DSIF-NELF. Nature 560, 601–606, doi:10.1038/s41586-018-0442-2 10.1038/s41586-018-0442-2 [pii] (2018). [PubMed: 30135580]
18. Vos SM et al. Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. Nature 560, 607–612, doi:10.1038/s41586-018-0440-4 (2018). [PubMed: 30135578]
19. Bieniasz PD, Grdina TA, Bogerd HP & Cullen BR Recruitment of cyclin T1/P-TEFb to an HIV type 1 long terminal repeat promoter proximal RNA target is both necessary and sufficient for full activation of transcription. Proc Natl Acad Sci U S A 96, 7791–7796, doi:10.1073/pnas.96.14.7791 (1999). [PubMed: 10393900]
20. Lis JT, Mason P, Peng J, Price DH & Werner J P-TEFb kinase recruitment and function at heat shock loci. Genes Dev 14, 792–803 (2000). [PubMed: 10766736]
21. 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, 265–276, doi:S0092–8674(05)00824-X [pii] 10.1016/j.cell.2005.08.019 (2005). [PubMed: 16239144]
22. Dominski Z, Yang XC, Purdy M, Wagner EJ & Marzluff WF 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 25, 1489–1500, doi:10.1128/MCB.25.4.1489-1500.2005 (2005). [PubMed: 15684398]
23. Ezzeddine N et al. A subset of Drosophila integrator proteins is essential for efficient U7 snRNA and spliceosomal snRNA 3'-end formation. Mol Cell Biol 31, 328–341, doi:10.1128/MCB.00943-10 (2011). [PubMed: 21078872]
24. Wu CW et al. RNA processing errors triggered by cadmium and integrator complex disruption are signals for environmental stress. BMC Biol 17, 56, doi:10.1186/s12915-019-0675-z (2019). [PubMed: 31311534]
25. Schmidt D et al. The Integrator complex regulates differential snRNA processing and fate of adult stem cells in the highly regenerative planarian Schmidtea mediterranea. PLoS Genet 14, e1007828, doi:10.1371/journal.pgen.1007828 (2018). [PubMed: 30557303]
26. Liu Y et al. snRNA 3' End Processing by a CPSF73-Containing Complex Essential for Development in Arabidopsis. PLoS Biol 14, e1002571, doi:10.1371/journal.pbio.1002571 (2016). [PubMed: 27780203]
27. 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, doi:10.1261/rna.035725.112 (2012). [PubMed: 23097424]
28. Pfleiderer MM & Galej WP Structure of the catalytic core of the Integrator complex. Mol Cell 81, 1246–1259 e1248, doi:S1097–2765(21)00005–8 [pii] 10.1016/j.molcel.2021.01.005 (2021). [PubMed: 33548203]
29. Sabath K et al. INTS10-INTS13-INTS14 form a functional module of Integrator that binds nucleic acids and the cleavage module. Nat Commun 11, 3422, doi:10.1038/s41467-020-17232-2 10.1038/s41467-020-17232-2 [pii] (2020). [PubMed: 32647223]
30. Barbieri E et al. Targeted Enhancer Activation by a Subunit of the Integrator Complex. Mol Cell 71, 103–116 e107, doi:S1097–2765(18)30406–4 [pii] 10.1016/j.molcel.2018.05.031 (2018). [PubMed: 30008316] Partition of Integrator into distinct functional modules is first proposed, with the identification of the enhancer module.
31. Vervoort SJ et al. The PP2A-Integrator-CDK9 axis fine-tunes transcription and can be targeted therapeutically in cancer. Cell, doi:S0092–8674(21)00502-X [pii] 10.1016/j.cell.2021.04.022 (2021).The phosphatase module of Integrator is identified as functionally opposing CDK9 activity at most protein coding genes.
32. Fianu I et al. Structural basis of Integrator-mediated transcription regulation. Science 374, 883–887, doi:doi:10.1126/science.abk0154 (2021). [PubMed: 34762484] This study presents Integrator's structure in association with reconstituted pausing Pol II, with an active INTS11 catalytic site.
33. Zheng H et al. Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. Science 370, doi:eabb5872 [pii] 10.1126/science.abb5872 370/6520/eabb5872 [pii]

(2020).The first cryo-EM analysis of Integrator reveals how core components assemble and identify the phosphatase module.

34. Huang KL et al. Integrator Recruits Protein Phosphatase 2A to Prevent Pause Release and Facilitate Transcription Termination. Mol Cell 80, 345–358 e349, doi:S1097-2765(20)30579-7 [pii] 10.1016/j.molcel.2020.08.016 (2020). [PubMed: 32966759]
35. Rengachari S, Schilbach S, Aibara S, Dienemann C & Cramer P Structure of the human Mediator-RNA polymerase II pre-initiation complex. Nature 594, 129–133, doi:10.1038/s41586-021-03555-7 (2021). [PubMed: 33902108]
36. Abdella R et al. Structure of the human Mediator-bound transcription preinitiation complex. Science 372, 52–56, doi:10.1126/science.abg3074 (2021). [PubMed: 33707221]
37. Sun Y et al. Structure of an active human histone pre-mRNA 3'-end processing machinery. Science 367, 700–703, doi:10.1126/science.aaz7758 (2020). [PubMed: 32029631]
38. Wu Y, Albrecht TR, Baillat D, Wagner EJ & Tong L Molecular basis for the interaction between Integrator subunits IntS9 and IntS11 and its functional importance. Proc Natl Acad Sci U S A 114, 4394–4399, doi:10.1073/pnas.1616605114 (2017). [PubMed: 28396433]
39. Albrecht TR et al. Integrator subunit 4 is a 'Symplekin-like' scaffold that associates with INTS9/11 to form the Integrator cleavage module. Nucleic Acids Res 46, 4241–4255, doi:10.1093/nar/gky100 (2018). [PubMed: 29471365]
40. Lambrecht C, Haesen D, Sents W, Ivanova E & Janssens V Structure, regulation, and pharmacological modulation of PP2A phosphatases. Methods Mol Biol 1053, 283–305, doi:10.1007/978-1-62703-562-0_17 (2013). [PubMed: 23860660]
41. Xu Y et al. Structure of the protein phosphatase 2A holoenzyme. Cell 127, 1239–1251, doi:S0092–8674(06)01537–6 [pii] 10.1016/j.cell.2006.11.033 (2006). [PubMed: 17174897]
42. Cho US & Xu W Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme. Nature 445, 53–57, doi:nature05351 [pii] 10.1038/nature05351 (2007). [PubMed: 17086192]
43. Seshacharyulu P, Pandey P, Datta K & Batra SK Phosphatase: PP2A structural importance, regulation and its aberrant expression in cancer. Cancer Letters 335, 9–18, doi:10.1016/j.canlet.2013.02.036 (2013). [PubMed: 23454242]
44. Ren W et al. Structural basis of SOSS1 complex assembly and recognition of ssDNA. Cell Rep 6, 982–991, doi:S2211–1247(14)00120-X [pii] 10.1016/j.celrep.2014.02.020 (2014). [PubMed: 24630995]
45. Arcus V OB-fold domains: a snapshot of the evolution of sequence, structure and function. Curr Opin Struct Biol 12, 794–801, doi:10.1016/s0959-440x(02)00392-5 (2002). [PubMed: 12504685]
46. Zhang F, Ma T & Yu X A core hSSB1-INTS complex participates in the DNA damage response. J Cell Sci 126, 4850–4855, doi:10.1242/jcs.132514 (2013). [PubMed: 23986477]
47. Jia Y et al. Crystal structure of the INTS3/INTS6 complex reveals the functional importance of INTS3 dimerization in DSB repair. Cell Discov 7, 66, doi:10.1038/s41421-021-00283-0 (2021). [PubMed: 34400606]
48. Li J et al. Structural basis for multifunctional roles of human Ints3 C-terminal domain. J Biol Chem 296, 100112, doi:10.1074/jbc.RA120.016393 (2021). [PubMed: 33434574]
49. Nojima T & Proudfoot NJ Mechanisms of lncRNA biogenesis as revealed by nascent transcriptomics. Nat Rev Mol Cell Biol, doi:10.1038/s41580-021-00447-6 (2022).
50. Hernandez N Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. J Biol Chem 276, 26733–26736, doi:10.1074/jbc.R100032200 (2001). [PubMed: 11390411]
51. Wilkinson ME, Charenton C & Nagai K RNA Splicing by the Spliceosome. Annu Rev Biochem 89, 359–388, doi:10.1146/annurev-biochem-091719-064225 (2020). [PubMed: 31794245]
52. Guiro J & Murphy S Regulation of expression of human RNA polymerase II-transcribed snRNA genes. Open Biol 7, doi:10.1098/rsob.170073 (2017).
53. Mandel CR et al. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. Nature 444, 953–956, doi:10.1038/nature05363 (2006). [PubMed: 17128255]
54. Egloff S, O'Reilly D & Murphy S Expression of human snRNA genes from beginning to end. Biochem Soc Trans 36, 590–594, doi:10.1042/BST0360590 (2008). [PubMed: 18631122]

55. Hernandez N & Weiner AM Formation of the 3' end of U1 snRNA requires compatible snRNA promoter elements. Cell 47, 249–258, doi:10.1016/0092-8674(86)90447-2 (1986). [PubMed: 3768956]
56. de Vegvar HE, Lund E & Dahlberg JE 3' end formation of U1 snRNA precursors is coupled to transcription from snRNA promoters. Cell 47, 259–266, doi:10.1016/0092-8674(86)90448-4 (1986). [PubMed: 3021336]
57. Hernandez N Formation of the 3' end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. EMBO J 4, 1827–1837 (1985). [PubMed: 2411548]
58. Lai F, Gardini A, Zhang A & Shiekhattar R Integrator mediates the biogenesis of enhancer RNAs. Nature 525, 399–403, doi:10.1038/nature14906 nature14906 [pii] (2015). [PubMed: 26308897]
59. Gomez-Orte E et al. Disruption of the Caenorhabditis elegans Integrator complex triggers a non-conventional transcriptional mechanism beyond snRNA genes. PLoS Genet 15, e1007981, doi:10.1371/journal.pgen.1007981 (2019). [PubMed: 30807579]
60. Davidson L, Francis L, Eaton JD & West S Integrator-Dependent and Allosteric/Intrinsic Mechanisms Ensure Efficient Termination of snRNA Transcription. Cell Rep 33, 108319, doi:10.1016/j.celrep.2020.108319 (2020). [PubMed: 33113359]
61. Egloff S et al. The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. The Journal of biological chemistry 285, 20564–20569, doi:10.1074/jbc.M110.132530 (2010). [PubMed: 20457598]
62. Ebmeier CC et al. Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications. Cell Rep 20, 1173–1186, doi:10.1016/j.celrep.2017.07.021 (2017). [PubMed: 28768201]
63. Yamamoto J et al. DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. Nature communications 5, 4263, doi:10.1038/ncomms5263 (2014).
64. Tatomer DC et al. The Integrator complex cleaves nascent mRNAs to attenuate transcription. Genes Dev 33, 1525–1538, doi:10.1101/gad.330167.119 gad.330167.119 [pii] (2019). [PubMed: 31530651]
65. O'Reilly D et al. Human snRNA genes use polyadenylation factors to promote efficient transcription termination. Nucleic Acids Res 42, 264–275, doi:10.1093/nar/gkt892 (2014). [PubMed: 24097444]
66. Baillat D, Gardini A, Cesaroni M & Shiekhattar R Requirement for SNAPC1 in transcriptional responsiveness to diverse extracellular signals. Mol Cell Biol 32, 4642–4650, doi:10.1128/MCB.00906-12 MCB.00906–12 [pii] (2012). [PubMed: 22966203]
67. Waldschmidt R, Wanandi I & Seifart KH Identification of transcription factors required for the expression of mammalian U6 genes in vitro. EMBO J 10, 2595–2603 (1991). [PubMed: 1868835]
68. Raha D et al. Close association of RNA polymerase II and many transcription factors with Pol III genes. Proc Natl Acad Sci U S A 107, 3639–3644, doi:10.1073/pnas.0911315106 (2010). [PubMed: 20139302]
69. Heinz S, Romanoski CE, Benner C & Glass CK The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol 16, 144–154, doi:10.1038/nrm3949 nrm3949 [pii] (2015). [PubMed: 25650801]
70. Adam RC et al. Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. Nature 521, 366–370, doi:10.1038/nature14289 (2015). [PubMed: 25799994]
71. Corces MR et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet 48, 1193–1203, doi:10.1038/ng.3646 (2016). [PubMed: 27526324]
72. Arner E et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science 347, 1010–1014, doi:10.1126/science.1259418 (2015). [PubMed: 25678556]
73. Field A & Adelman K Evaluating Enhancer Function and Transcription. Annu Rev Biochem 89, 213–234, doi:10.1146/annurev-biochem-011420-095916 (2020). [PubMed: 32197056]
74. Schoenfelder S & Fraser P Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet 20, 437–455, doi:10.1038/s41576-019-0128-0 (2019). [PubMed: 31086298]

75. Andersson R et al. An atlas of active enhancers across human cell types and tissues. Nature 507, 455–461, doi:10.1038/nature12787 nature12787 [pii] (2014). [PubMed: 24670763]
76. Kim TK et al. Widespread transcription at neuronal activity-regulated enhancers. Nature 465, 182–187, doi:10.1038/nature09033 nature09033 [pii] (2010). [PubMed: 20393465]
77. Statello L, Guo CJ, Chen LL & Huarte M Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol 22, 96–118, doi:10.1038/s41580-020-00315-9 (2021). [PubMed: 33353982]
78. De Santa F et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol 8, e1000384, doi:10.1371/journal.pbio.1000384 (2010). [PubMed: 20485488]
79. Gil N & Ulitsky I Production of Spliced Long Noncoding RNAs Specifies Regions with Increased Enhancer Activity. Cell Syst 7, 537–547 e533, doi:10.1016/j.cels.2018.10.009 (2018). [PubMed: 30447999]
80. van den Berg DLC et al. Nipbl Interacts with Zfp609 and the Integrator Complex to Regulate Cortical Neuron Migration. Neuron 93, 348–361, doi:10.1016/j.neuron.2016.11.047 (2017). [PubMed: 28041881]
81. Gurumurthy A et al. Super-enhancer mediated regulation of adult beta-globin gene expression: the role of eRNA and Integrator. Nucleic Acids Res 49, 1383–1396, doi:10.1093/nar/gkab002 (2021). [PubMed: 33476375]
82. Nojima T et al. Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. Mol Cell 72, 970–984 e977, doi:10.1016/j.molcel.2018.10.011 (2018). [PubMed: 30449723]
83. Cazalla D, Xie M & Steitz JA A primate herpesvirus uses the integrator complex to generate viral microRNAs. Molecular cell 43, 982–992, doi:10.1016/j.molcel.2011.07.025 (2011). [PubMed: 21925386]
84. Xie M et al. The host Integrator complex acts in transcription-independent maturation of herpesvirus microRNA 3' ends. Genes Dev 29, 1552–1564, doi:10.1101/gad.266973.115 (2015). [PubMed: 26220997]
85. Kirstein N et al. The Integrator complex regulates microRNA abundance through RISC loading. bioRxiv, 2021.2009.2021.461113, doi:10.1101/2021.09.21.461113 (2021).
86. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D & Zamore PD PIWI-interacting RNAs: small RNAs with big functions. Nat Rev Genet 20, 89–108, doi:10.1038/s41576-018-0073-3 (2019). [PubMed: 30446728]
87. Beltran T, Pahita E, Ghosh S, Lenhard B & Sarkies P Integrator is recruited to promoter-proximally paused RNA Pol II to generate Caenorhabditis elegans piRNA precursors. EMBO J 40, e105564, doi:10.15252/embj.2020105564 (2021). [PubMed: 33340372]
88. Berkyurek AC et al. The RNA polymerase II subunit RPB-9 recruits the integrator complex to terminate Caenorhabditis elegans piRNA transcription. EMBO J 40, e105565, doi:10.15252/embj.2020105565 (2021). [PubMed: 33533030]
89. Czech B, Preall JB, McGinn J & Hannon GJ A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. Mol Cell 50, 749–761, doi:10.1016/j.molcel.2013.04.007 (2013). [PubMed: 23665227]
90. Handler D et al. The genetic makeup of the Drosophila piRNA pathway. Mol Cell 50, 762–777, doi:10.1016/j.molcel.2013.04.031 (2013). [PubMed: 23665231]
91. Shay JW & Wright WE Telomeres and telomerase: three decades of progress. Nat Rev Genet 20, 299–309, doi:10.1038/s41576-019-0099-1 (2019). [PubMed: 30760854]
92. Rubtsova MP et al. Integrator is a key component of human telomerase RNA biogenesis. Sci Rep 9, 1701, doi:10.1038/s41598-018-38297-6 (2019). [PubMed: 30737432]
93. Clemson CM et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell 33, 717–726, doi:10.1016/j.molcel.2009.01.026 (2009). [PubMed: 19217333]
94. Barra J et al. Integrator restrains paraspeckles assembly by promoting isoform switching of the lncRNA NEAT1. Sci Adv 6, eaaz9072, doi:10.1126/sciadv.aaz9072 (2020).

95. Naganuma T et al. Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. EMBO J 31, 4020–4034, doi:10.1038/emboj.2012.251 (2012). [PubMed: 22960638]

96. Rasmussen EB & Lis JT In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc Natl Acad Sci U S A 90, 7923–7927 (1993). [PubMed: 8367444]

97. Core LJ & Lis JT Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science 319, 1791–1792, doi:10.1126/science.1150843 319/5871/1791 [pii] (2008). [PubMed: 18369138]

98. Adelman K & Lis JT Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nature reviews. Genetics 13, 720–731, doi:10.1038/nrg3293 (2012).

99. Chen FX et al. PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. Cell 162, 1003–1015, doi:10.1016/j.cell.2015.07.042 S0092–8674(15)00958–7 [pii] (2015). [PubMed: 26279188]

100. Chen FX et al. PAF1 regulation of promoter-proximal pause release via enhancer activation. Science 357, 1294–1298, doi:10.1126/science.aan3269 science.aan3269 [pii] (2017). [PubMed: 28860207]

101. Core LJ et al. Defining the status of RNA polymerase at promoters. Cell reports 2, 1025–1035, doi:10.1016/j.celrep.2012.08.034 (2012). [PubMed: 23062713]

102. Liu J, Wu X, Zhang H, Pfeifer GP & Lu Q Dynamics of RNA Polymerase II Pausing and Bivalent Histone H3 Methylation during Neuronal Differentiation in Brain Development. Cell Rep 20, 1307–1318, doi:S2211–1247(17)31025–2 [pii] 10.1016/j.celrep.2017.07.046 (2017). [PubMed: 28793256]

103. Gaertner B & Zeitlinger J RNA polymerase II pausing during development. Development 141, 1179–1183, doi:10.1242/dev.088492 141/6/1179 [pii] (2014). [PubMed: 24595285]

104. Danko CG et al. Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. Mol Cell 50, 212–222, doi:10.1016/j.molcel.2013.02.015 S1097–2765(13)00171–8 [pii] (2013). [PubMed: 23523369]

105. Galbraith MD et al. HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. Cell 153, 1327–1339, doi:10.1016/j.cell.2013.04.048 S0092–8674(13)00524–2 [pii] (2013). [PubMed: 23746844]

106. Nilson KA et al. Oxidative stress rapidly stabilizes promoter-proximal paused Pol II across the human genome. Nucleic Acids Res 45, 11088–11105, doi:10.1093/nar/gkx724 4084663 [pii] (2017). [PubMed: 28977633]

107. Andrulis ED, Guzman E, Doring P, Werner J & Lis JT High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. Genes Dev 14, 2635–2649, doi:10.1101/gad.844200 (2000). [PubMed: 11040217]

108. Gardini A et al. Integrator regulates transcriptional initiation and pause release following activation. Mol Cell 56, 128–139, doi:10.1016/j.molcel.2014.08.004 (2014). [PubMed: 25201415] The first genome-wide analysis of Integrator occupancy reveals diffuse binding at Pol II genes and requirement for stimulus-dependent transcriptional activation.

109. Beckedorff F et al. The Human Integrator Complex Facilitates Transcriptional Elongation by Endonucleolytic Cleavage of Nascent Transcripts. Cell Rep 32, 107917, doi:S2211–1247(20)30898–6 [pii] 10.1016/j.celrep.2020.107917 (2020). [PubMed: 32697989] This study proposes an RNA cleavage-dependent mechanism that promotes productive Pol II elongation in mammals.

110. Stadelmayer B et al. Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. Nature communications 5, 5531, doi:10.1038/ncomms6531 (2014).

111. Elrod ND et al. The Integrator Complex Attenuates Promoter-Proximal Transcription at Protein-Coding Genes. Mol Cell 76, 738–752 e737, doi:S1097–2765(19)30812–3 [pii] 10.1016/j.molcel.2019.10.034 (2019). [PubMed: 31809743] This study implicates the Integrator complex in transcription attenuation of protein coding genes, via the endonucleolytic module.

112. Aoi Y et al. NELF Regulates a Promoter-Proximal Step Distinct from RNA Pol II Pause-Release. Mol Cell 78, 261–274 e265, doi:10.1016/j.molcel.2020.02.014 (2020). [PubMed: 32155413]

113. Shersher E et al. NACK and INTEGRATOR act coordinately to activate Notch-mediated transcription in tumorigenesis. Cell Commun Signal 19, 96, doi:10.1186/s12964-021-00776-1 (2021). [PubMed: 34551776]
114. Kamieniarz-Gdula K et al. Selective Roles of Vertebrate PCF11 in Premature and Full-Length Transcript Termination. Mol Cell 74, 158–172 e159, doi:10.1016/j.molcel.2019.01.027 (2019). [PubMed: 30819644]
115. Kamieniarz-Gdula K & Proudfoot NJ Transcriptional Control by Premature Termination: A Forgotten Mechanism. Trends Genet 35, 553–564, doi:10.1016/j.tig.2019.05.005 (2019). [PubMed: 31213387]
116. Lykke-Andersen S et al. Integrator is a genome-wide attenuator of non-productive transcription. Mol Cell 81, 514–529 e516, doi:S1097–2765(20)30906–0 [pii] 10.1016/j.molcel.2020.12.014 (2021). [PubMed: 33385327]
117. Skaar JR et al. The Integrator complex controls the termination of transcription at diverse classes of gene targets. Cell research 25, 288–305, doi:10.1038/cr.2015.19 (2015). [PubMed: 25675981]
118. Dasilva LF et al. Integrator enforces the fidelity of transcriptional termination at protein-coding genes. Science Advances 7, eabe3393, doi:doi:10.1126/sciadv.abe3393 (2021).
119. Rosa-Mercado NA et al. Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression. Mol Cell 81, 502–513 e504, doi:10.1016/j.molcel.2020.12.002 (2021). [PubMed: 33400923]
120. Cossa G, Parua PK, Eilers M & Fisher RP Protein phosphatases in the RNAPII transcription cycle: erasers, sculptors, gatekeepers, and potential drug targets. Genes Dev 35, 658–676, doi:10.1101/gad.348315.121 (2021). [PubMed: 33888562]
121. Sents W, Ivanova E, Lambrecht C, Haesen D & Janssens V The biogenesis of active protein phosphatase 2A holoenzymes: a tightly regulated process creating phosphatase specificity. FEBS J 280, 644–661, doi:10.1111/j.1742-4658.2012.08579.x (2013). [PubMed: 22443683]
122. Malovannaya A et al. Analysis of the human endogenous coregulator complexome. Cell 145, 787–799, doi:10.1016/j.cell.2011.05.006 (2011). [PubMed: 21620140]
123. Malovannaya A et al. Streamlined analysis schema for high-throughput identification of endogenous protein complexes. Proc Natl Acad Sci U S A 107, 2431–2436, doi:10.1073/pnas.0912599106 (2010). [PubMed: 20133760]
124. Core L & Adelman K Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev 33, 960–982, doi:10.1101/gad.325142.119 (2019). [PubMed: 31123063]
125. Jeronimo C, Collin P & Robert F The RNA Polymerase II CTD: The Increasing Complexity of a Low-Complexity Protein Domain. J Mol Biol 428, 2607–2622, doi:10.1016/j.jmb.2016.02.006 (2016). [PubMed: 26876604]
126. Galbraith MD, Bender H & Espinosa JM Therapeutic targeting of transcriptional cyclin-dependent kinases. Transcription 10, 118–136, doi:10.1080/21541264.2018.1539615 (2019). [PubMed: 30409083]
127. Parua PK & Fisher RP Dissecting the Pol II transcription cycle and derailing cancer with CDK inhibitors. Nat Chem Biol 16, 716–724, doi:10.1038/s41589-020-0563-4 10.1038/s41589-020-0563-4 [pii] (2020). [PubMed: 32572259]
128. Luo Z, Lin C & Shilatifard A The super elongation complex (SEC) family in transcriptional control. Nat Rev Mol Cell Biol 13, 543–547, doi:10.1038/nrm3417 nrm3417 [pii] (2012). [PubMed: 22895430]
129. Hu S et al. SPT5 stabilizes RNA polymerase II, orchestrates transcription cycles, and maintains the enhancer landscape. Mol Cell 81, 4425–4439 e4426, doi:10.1016/j.molcel.2021.08.029 (2021). [PubMed: 34534457]
130. Parua PK, Kalan S, Benjamin B, Sanso M & Fisher RP Distinct Cdk9-phosphatase switches act at the beginning and end of elongation by RNA polymerase II. Nat Commun 11, 4338, doi:10.1038/s41467-020-18173-6 (2020). [PubMed: 32859893]
131. Parua PK et al. A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. Nature 558, 460–464, doi:10.1038/s41586-018-0214-z (2018). [PubMed: 29899453]

132. Mascibroda LG et al. *INTS13* Mutations Causing a Developmental Ciliopathy Disrupt Integrator Complex Assembly. bioRxiv, 2020.2007.2020.209130, doi:10.1101/2020.07.20.209130 (2020).
133. Skaar JR et al. INTS3 controls the hSSB1-mediated DNA damage response. J Cell Biol 187, 25–32, doi:10.1083/jcb.200907026 (2009). [PubMed: 19786574]
134. Li Y et al. HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. J Biol Chem 284, 23525–23531, doi:10.1074/jbc.C109.039586 (2009). [PubMed: 19605351]
135. Huang J, Gong Z, Ghosal G & Chen J SOSS complexes participate in the maintenance of genomic stability. Mol Cell 35, 384–393, doi:10.1016/j.molcel.2009.06.011 (2009). [PubMed: 19683501]
136. Byrne BM & Oakley GG Replication protein A, the laxative that keeps DNA regular: The importance of RPA phosphorylation in maintaining genome stability. Semin Cell Dev Biol 86, 112–120, doi:10.1016/j.semcdb.2018.04.005 (2019). [PubMed: 29665433]
137. Richard DJ et al. Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. Nature 453, 677–681, doi:10.1038/nature06883 (2008). [PubMed: 18449195]
138. Richard DJ et al. hSSB1 rapidly binds at the sites of DNA double-strand breaks and is required for the efficient recruitment of the MRN complex. Nucleic Acids Res 39, 1692–1702, doi:10.1093/nar/gkq1098 (2011). [PubMed: 21051358]
139. Richard DJ et al. hSSB1 interacts directly with the MRN complex stimulating its recruitment to DNA double-strand breaks and its endo-nuclease activity. Nucleic Acids Res 39, 3643–3651, doi:10.1093/nar/gkq1340 (2011). [PubMed: 21227926]
140. Oegema R et al. Human mutations in integrator complex subunits link transcriptome integrity to brain development. PLoS Genet 13, e1006809, doi:10.1371/journal.pgen.1006809 PGENETICS-D-16-01999 [pii] (2017). [PubMed: 28542170] First report that recessive mutations of Integrator subunits are linked to severe developmental disorders.
141. Yoshimi A et al. Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis. Nature 574, 273–277, doi:10.1038/s41586-019-1618-0 (2019). [PubMed: 31578525] This study demonstrates that loss of a subunit of Integrator, INTS3, functions as a driver of leukemogenesis.
142. Kheirallah AK, de Moor CH, Faiz A, Sayers I & Hall IP Lung function associated gene Integrator Complex subunit 12 regulates protein synthesis pathways. BMC Genomics 18, 248, doi:10.1186/s12864-017-3628-3 (2017). [PubMed: 28335732]
143. Hata T & Nakayama M Targeted disruption of the murine large nuclear KIAA1440/Ints1 protein causes growth arrest in early blastocyst stage embryos and eventual apoptotic cell death. Biochim Biophys Acta 1773, 1039–1051, doi:10.1016/j.bbamcr.2007.04.010 (2007). [PubMed: 17544522]
144. Kapp LD, Abrams EW, Marlow FL & Mullins MC The integrator complex subunit 6 (Ints6) confines the dorsal organizer in vertebrate embryogenesis. PLoS Genet 9, e1003822, doi:10.1371/journal.pgen.1003822 (2013). [PubMed: 24204286]
145. Huang H et al. The integrator complex subunit 11 is involved in the post-diapaused embryonic development and stress response of Artemia sinica. Gene 741, 144548, doi:10.1016/j.gene.2020.144548 (2020). [PubMed: 32165292]
146. Tao S, Cai Y & Sampath K The Integrator subunits function in hematopoiesis by modulating Smad/BMP signaling. Development 136, 2757–2765, doi:10.1242/dev.034959 (2009). [PubMed: 19605500]
147. Zhang P et al. INTS11 regulates hematopoiesis by promoting PRC2 function. Sci Adv 7, eabh1684, doi:10.1126/sciadv.abh1684 (2021).
148. Otani Y et al. Integrator complex plays an essential role in adipose differentiation. Biochemical and biophysical research communications 434, 197–202, doi:10.1016/j.bbrc.2013.03.029 (2013). [PubMed: 23523797]
149. Zhang Y et al. The Integrator Complex Prevents Dedifferentiation of Intermediate Neural Progenitors back into Neural Stem Cells. Cell Rep 27, 987–996 e983, doi:10.1016/j.celrep.2019.03.089 (2019). [PubMed: 31018143]

150. Krall M et al. Biallelic sequence variants in INTS1 in patients with developmental delays, cataracts, and craniofacial anomalies. Eur J Hum Genet 27, 582–593, doi:10.1038/s41431-018-0298-9 (2019). [PubMed: 30622326]
151. Zhang X et al. Biallelic INTS1 Mutations Cause a Rare Neurodevelopmental Disorder in Two Chinese Siblings. J Mol Neurosci 70, 1–8, doi:10.1007/s12031-019-01393-x (2020). [PubMed: 31428919]
152. Bacon C & Rappold GA The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. Hum Genet 131, 1687–1698, doi:10.1007/s00439-012-1193-z (2012). [PubMed: 22736078]
153. Lenaerts L et al. The broad phenotypic spectrum of PPP2R1A-related neurodevelopmental disorders correlates with the degree of biochemical dysfunction. Genet Med 23, 352–362, doi:10.1038/s41436-020-00981-2 10.1038/s41436020-009812 [pii] (2021). [PubMed: 33106617]
154. Zhang Y et al. A De Novo Variant Identified in the PPP2R1A Gene in an Infant Induces Neurodevelopmental Abnormalities. Neurosci Bull 36, 179–182, doi:10.1007/s12264-019-00430-4 (2020). [PubMed: 31531803]
155. Wallace A, Caruso P & Karaa A A Newborn with Severe Ventriculomegaly: Expanding the PPP2R1A Gene Mutation Phenotype. J Pediatr Genet 8, 240–243, doi:10.1055/s-0039-1692414 (2019). [PubMed: 31687265]
156. Houge G et al. B56delta-related protein phosphatase 2A dysfunction identified in patients with intellectual disability. J Clin Invest 125, 3051–3062, doi:10.1172/JCI79860 (2015). [PubMed: 26168268]
157. Deciphering Developmental Disorders S Large-scale discovery of novel genetic causes of developmental disorders. Nature 519, 223–228, doi:10.1038/nature14135 (2015). [PubMed: 25533962]
158. Tilley FC et al. Disruption of pathways regulated by Integrator complex in Galloway-Mowat syndrome due to WDR73 mutations. Sci Rep 11, 5388, doi:10.1038/s41598-021-84472-7 (2021). [PubMed: 33686175]
159. Wheway G, Nazlamova L & Hancock JT Signaling through the Primary Cilium. Front Cell Dev Biol 6, 8, doi:10.3389/fcell.2018.00008 (2018). [PubMed: 29473038]
160. Jodoin JN et al. Nuclear-localized Asunder regulates cytoplasmic dynein localization via its role in the integrator complex. Mol Biol Cell 24, 2954–2965, doi:10.1091/mbc.E13-05-0254 (2013). [PubMed: 23904267]
161. Jodoin JN et al. The snRNA-processing complex, Integrator, is required for ciliogenesis and dynein recruitment to the nuclear envelope via distinct mechanisms. Biol Open 2, 1390–1396, doi:10.1242/bio.20136981 (2013). [PubMed: 24285713]
162. Mittal P & Roberts CWM The SWI/SNF complex in cancer - biology, biomarkers and therapy. Nat Rev Clin Oncol 17, 435–448, doi:10.1038/s41571-020-0357-3 (2020). [PubMed: 32303701]
163. Federico A et al. Pan-Cancer Mutational and Transcriptional Analysis of the Integrator Complex. Int J Mol Sci 18, doi:10.3390/ijms18050936 (2017).
164. Van den Eynden J, Basu S & Larsson E Somatic Mutation Patterns in Hemizygous Genomic Regions Unveil Purifying Selection during Tumor Evolution. PLoS Genet 12, e1006506, doi:10.1371/journal.pgen.1006506 (2016). [PubMed: 28027311]
165. Yue J et al. Integrator orchestrates RAS/ERK1/2 signaling transcriptional programs. Genes Dev 31, 1809–1820, doi:10.1101/gad.301697.117 31/17/1809 [pii] (2017). [PubMed: 28982763]
166. Tong H et al. INTS8 accelerates the epithelial-to-mesenchymal transition in hepatocellular carcinoma by upregulating the TGF-beta signaling pathway. Cancer Manag Res 11, 1869–1879, doi:10.2147/CMAR.S184392 (2019). [PubMed: 30881114]
167. Inagaki Y et al. CREB3L4, INTS3, and SNAPAP are targets for the 1q21 amplicon frequently detected in hepatocellular carcinoma. Cancer Genet Cytogenet 180, 30–36, doi:10.1016/j.cancergencyto.2007.09.013 (2008). [PubMed: 18068530]
168. Wieland I et al. Isolation of DICE1: a gene frequently affected by LOH and downregulated in lung carcinomas. Oncogene 18, 4530–4537, doi:10.1038/sj.onc.1202806 (1999). [PubMed: 10467397]

169. Filleur S 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 9, 28, doi:10.1186/1475-2867-9-28 (2009). [PubMed: 19906297]

170. Li J et al. Bioinformatics analysis of gene expression profiles in childhood B-precursor acute lymphoblastic leukemia. Hematology 20, 377–383, doi:10.1179/1607845414Y.0000000214 (2015). [PubMed: 25431969]

171. Ropke A et al. Promoter CpG hypermethylation and downregulation of DICE1 expression in prostate cancer. Oncogene 24, 6667–6675, doi:10.1038/sj.onc.1208824 (2005). [PubMed: 16007164]

172. Perrotti D & Neviani P Protein phosphatase 2A: a target for anticancer therapy. Lancet Oncol 14, e229–238, doi:10.1016/S1470-2045(12)70558-2 (2013). [PubMed: 23639323]

173. O'Connor CM et al. Inactivation of PP2A by a recurrent mutation drives resistance to MEK inhibitors. Oncogene 39, 703–717, doi:10.1038/s41388-019-1012-2 10.1038/s41388-019-1012-2 [pii] (2020). [PubMed: 31541192]

174. Taylor SE et al. The Highly Recurrent PP2A Aalpha-Subunit Mutation P179R Alters Protein Structure and Impairs PP2A Enzyme Function to Promote Endometrial Tumorigenesis. Cancer Res 79, 4242–4257, doi:10.1158/0008-5472.CAN-19-0218 0008–5472.CAN-19–0218 [pii] (2019). [PubMed: 31142515]

175. Cherniack AD et al. Integrated Molecular Characterization of Uterine Carcinosarcoma. Cancer Cell 31, 411–423, doi:S1535–6108(17)30053–3 [pii] 10.1016/j.ccell.2017.02.010 (2017). [PubMed: 28292439]

176. Haesen D et al. Recurrent PPP2R1A Mutations in Uterine Cancer Act through a Dominant-Negative Mechanism to Promote Malignant Cell Growth. Cancer Res 76, 5719–5731, doi:0008–5472.CAN-15–3342 [pii] 10.1158/0008-5472.CAN-15-3342 (2016). [PubMed: 27485451]

177. Shih Ie M et al. Somatic mutations of PPP2R1A in ovarian and uterine carcinomas. Am J Pathol 178, 1442–1447, doi:10.1016/j.ajpath.2011.01.009 S0002–9440(11)00060–5 [pii] (2011). [PubMed: 21435433]

178. Bockelman C et al. Prognostic role of CIP2A expression in serous ovarian cancer. Br J Cancer 105, 989–995, doi:10.1038/bjc.2011.346 (2011). [PubMed: 21897396]

179. Leonard D et al. Selective PP2A Enhancement through Biased Heterotrimer Stabilization. Cell 181, 688–701 e616, doi:10.1016/j.cell.2020.03.038 (2020). [PubMed: 32315618]

180. Neviani P et al. PP2A-activating drugs selectively eradicate TKI-resistant chronic myeloid leukemic stem cells. J Clin Invest 123, 4144–4157, doi:10.1172/JCI68951 (2013). [PubMed: 23999433]

181. Kastrinsky DB et al. Reengineered tricyclic anti-cancer agents. Bioorg Med Chem 23, 6528–6534, doi:10.1016/j.bmc.2015.07.007 (2015). [PubMed: 26372073]

182. Harlen KM & Churchman LS The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nat Rev Mol Cell Biol 18, 263–273, doi:10.1038/nrm.2017.10 (2017). [PubMed: 28248323]

183. Schuller R et al. Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Mol Cell 61, 305–314, doi:10.1016/j.molcel.2015.12.003 (2016). [PubMed: 26799765]

Box 1.

The C-terminal domain of RNA polymerase II

RPB1, the catalytic subunit of RNA polymerase II (Pol II) has a large, disordered region known as the C-terminal domain (CTD)[182]. The CTD is a low-complexity domain of 378 aa, essentially composed of tandem heptad repeats of Y-S-P-T-S-P-S. Although the heptad is highly conserved from yeast to humans, the number of repeats varies from 26 in *Saccharomyces Cerevisiae* to 52 in human. The Pol II CTD is crucial for transcription regulation, through phosphorylation of its heptad repeats by multiple kinases from the PIC assembly steps to termination. The combination of phosphorylated residues is thought to function as a biochemical code that enables Pol II to recruit a broad range of protein complexes, including Mediator, Integrator, the capping machinery, the spliceosome and cleavage and polyadenylation specificity factor[182]. All the heptads can be phosphorylated in vivo, each carrying one or two modified residues on average[183]. The most targeted residues are Ser2 and Ser5 and their phosphorylation is associated with Pol II pause release into productive elongation and with transcription initiation, respectively[183]. Ser7 phosphorylation is commonly found at Pol II molecules across the bodies of coding and non-coding loci and it may have a role in 3’-end formation of U small nuclear RNAs[61]. As a disordered, low-complexity protein domain, the CTD has not been structurally characterized, with the exception of few repeats that have been modeled when bound to the Mediator and the Integrator complexes[32,36]. Notably, Integrator has robust affinity for the CTD in both its unphosphorylated and phosphorylated forms[21,62] and the Int–PP2A module is capable of directly dephosphorylating Ser residues[31,33]. Recent structural data also suggest that phosphorylated Tyr1 residues may enhance the affinity of Integrator for the CTD[32].

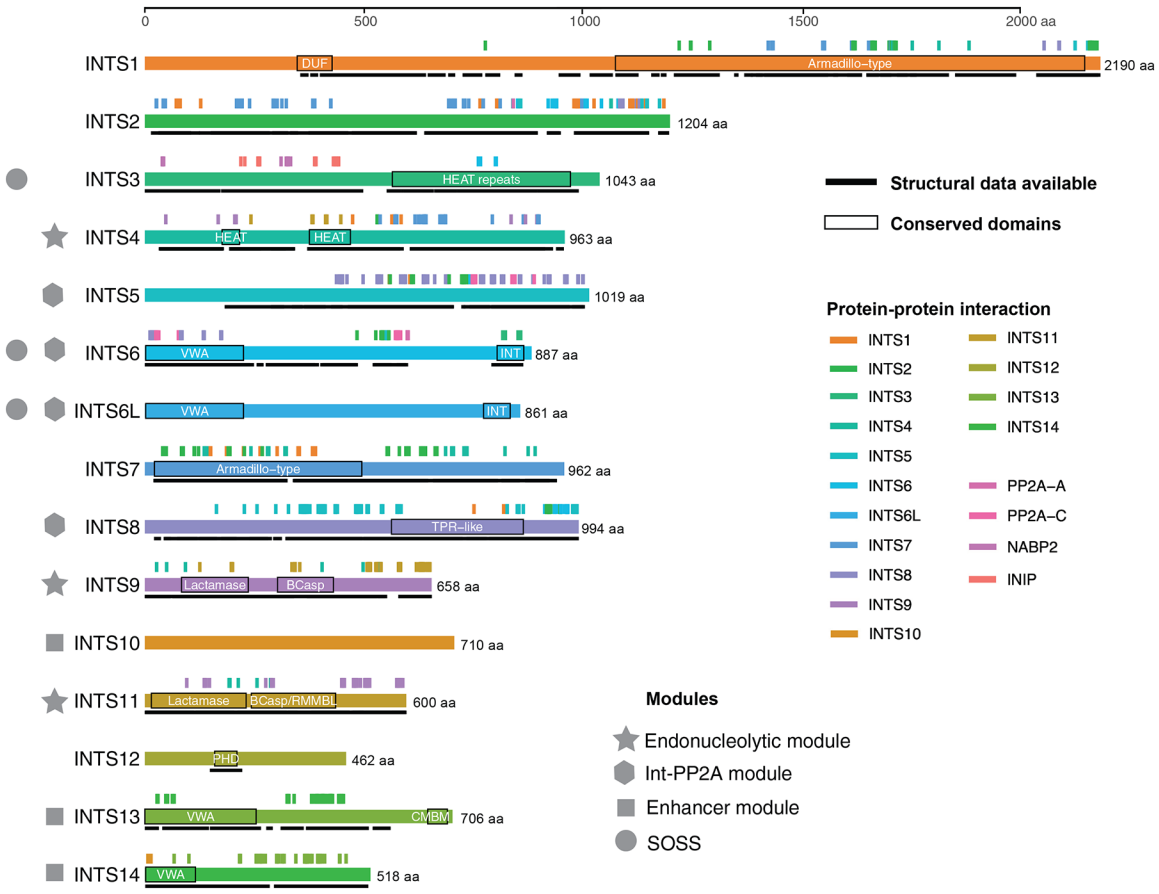

Figure 1. Subunits of the mammalian Integrator complex.

The 15 known subunits of the mammalian Integrator complex are depicted to scale. Both INTS6 and its mutually exclusive paralog INTS6L are shown. All annotated protein motifs and domains are named and boxed. The availability of published structural data (cryo-EM or X-ray crystallography) is marked by an underline. Colored bars on top of each subunit diagram denotes a mapped interaction surface with another member of Integrator, protein phosphatase 2A (PP2A) or sensor of ssDNA (SOSS) complex, according to the specific color code. Lastly, a symbol on the left end indicates whether the corresponding subunit is a component of the endonucleolytic, Integrator–PP2A (Int-PP2A) or enhancer modules, or whether it is found in the SOSS DNA repair complex.

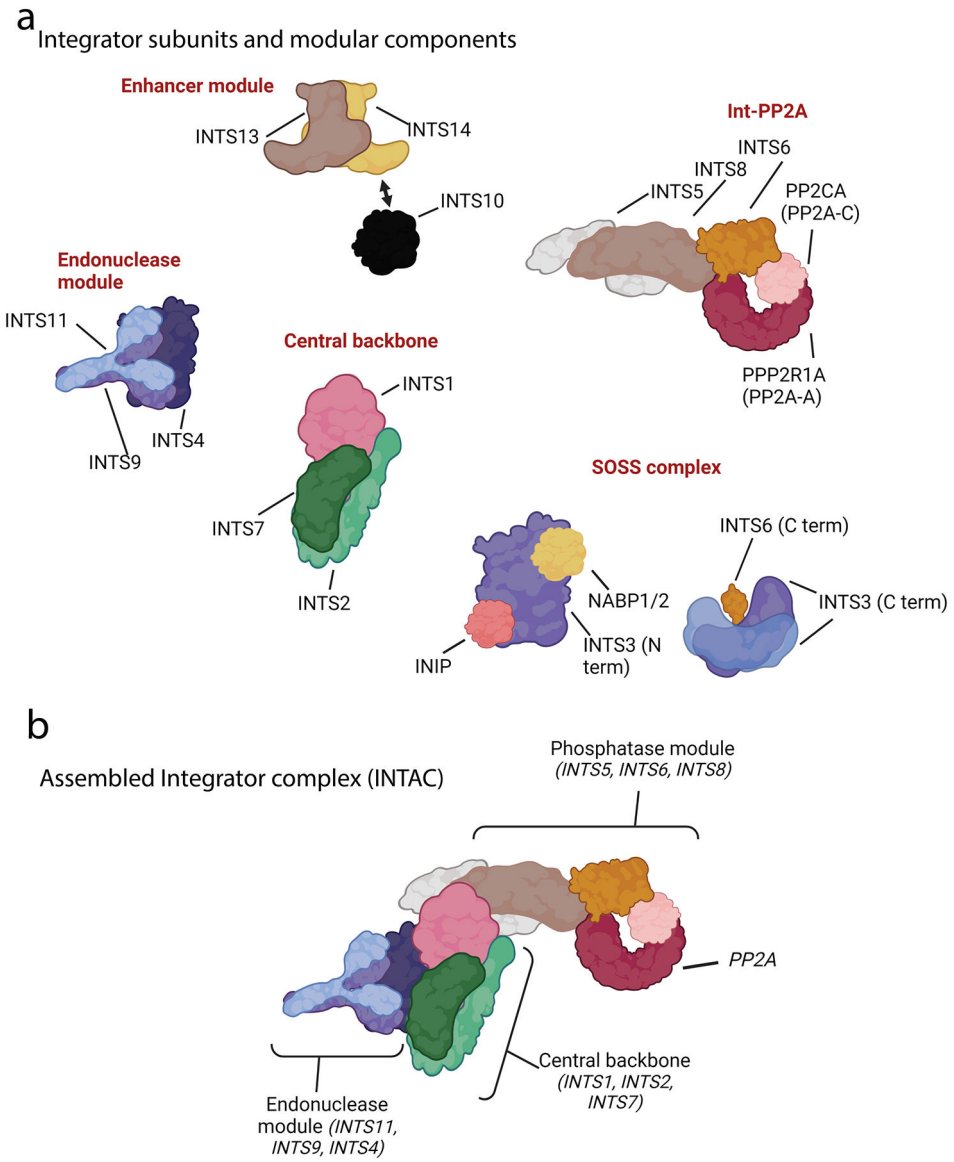

Figure 2. Structural features and assembly of Integrator.

A) The functional modules of Integrator are biochemically and structurally distinct. The enhancer module comprises INTS13 and INTS14, which stabilize one another when forming a heterodimer. INTS10 biochemically associates with the INTS13–INTS14 heterodimer, although structural information on INTS10 has not yet been obtained. The Int– serine/threonine-protein phosphatase 2A (PP2A) module comprises the PP2A heterodimer, which includes the scaffold subunit PP2A-A the smaller catalytic subunit PP2A-C, assembled on INTS5 and INTS8. INTS6 further bridges the PP2A dimer to INTS5–INTS8A. The endonuclease module can be reconstituted as a stable trimer of INTS11 (catalytic subunit), INTS9 (catalytic, inactive) and INTS4 (scaffold subunit). The sensor of ssDNA (SOSS) complex, which is implicated in DNA repair, is assembled around INTS3. Although a structure of the complete SOSS complex is not available, the N-terminal half of INTS3 has been crystallized with the small subunits of SOSS, nucleic acid-binding protein 1 (NABP1)

or NABP2, and INTS3 and NABP interacting protein (INIP). The remaining C-terminal moiety of INTS3 has been independently crystallized as a homodimer, in association with a small C-terminal tail of INTS6.

B) The structure of the fully assembled Integrator complex with PP2A (INTAC) reveals a central backbone (INTS1, INTS2 and INTS7) accommodating the two catalytic modules (endonuclease and phosphatase) on opposite sides.

Protein Data Bank entries with structures of Integrator: 7CUN, 7PKS, 7BFQ, 7BFP, 6SN1, 7BV7, 6WLG, 4OWW.

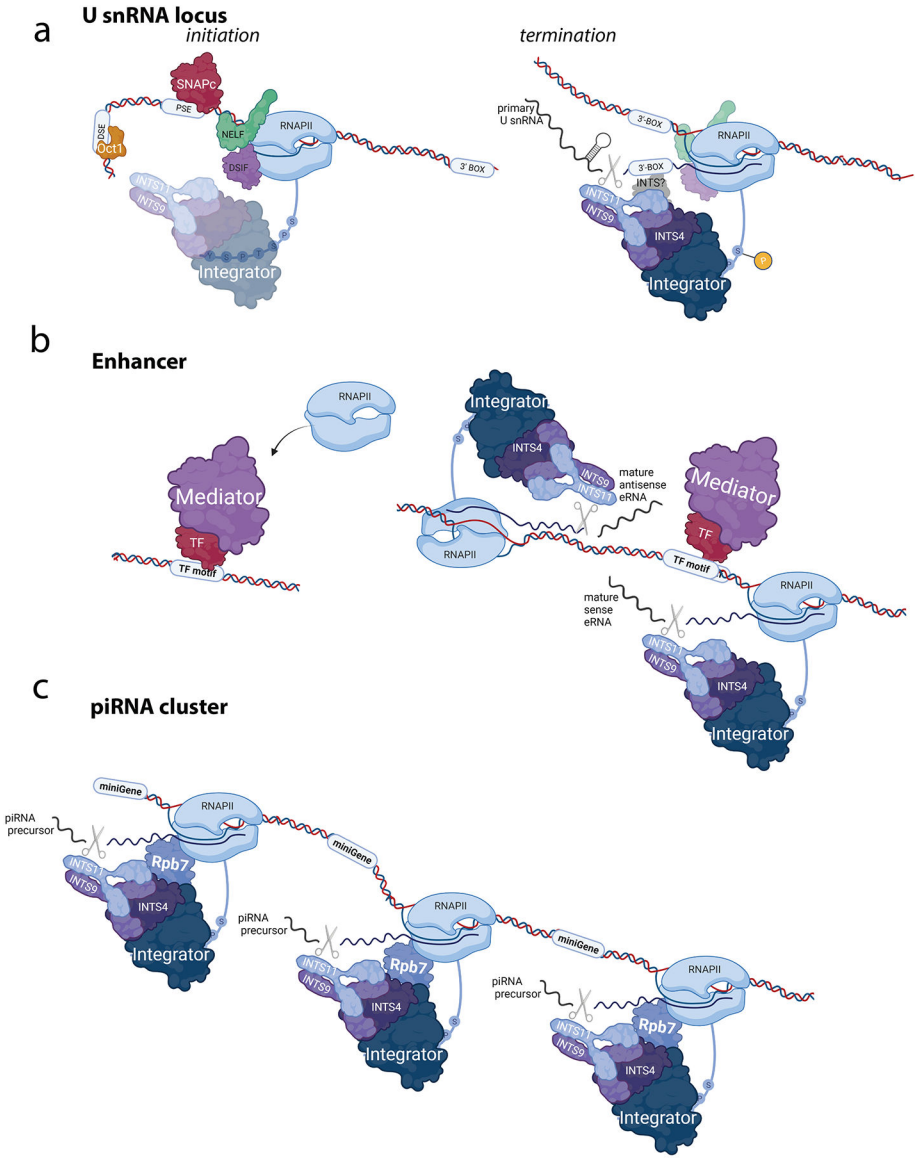

Figure 3. Integrator terminates transcription of non-coding RNAs.

A) U small nuclear RNA (U snRNA) loci have distinctive promoter elements termed distal sequence element (DSE) and proximal sequence element (PSE), which bind the OCT1 (POU2F1) transcription factor and the small nuclear RNA-activating protein complex (SNAPc). OCT1 and SNAPc contribute to the recruitment of a transcription initiation-competent RNA polymerase II (Pol II) holoenzyme including DRB-sensitivity inducing factor (DSIF), negative elongation factor (NELF) and the Integrator complex. Shortly after transcribing through the 3' box, which is a highly conserved motif at the termination site of all U snRNAs, Integrator cleaves the nascent small RNA, triggered by phosphorylation (P) of Ser7 of the C-terminal domain (CTD) of Pol II's largest subunit (RBP1). A 3' stem-loop in the precursor U snRNA and recognition of the ensuing 3' box RNA sequence by a set of Integrator accessory subunits (INTS?) may support an efficient cleavage process.

B) Enhancer loci are activated by sequence-specific transcription factors that recruit the co-activator Mediator complex. Upon Mediator recruitment and assembly of the transcription pre-initiation complex (not shown), bi-directional transcription of the enhancer locus occurs, producing long (>200bp) sense and antisense transcripts that are called enhancer RNAs (eRNAs). Both sense-transcribing and antisense-transcribing Pol II holoenzymes recruit the Integrator complex to terminate transcription and release eRNAs without eliciting their polyadenylation.

C) *Caenorhabditis elegans* PIWI-interacting RNAs (piRNAs) are generally encoded by mini-genes grouped into large clusters. A 20–40bp piRNA precursor is cleaved co-transcriptionally by the Integrator complex, which is recruited by the Pol II CTD and by the Pol II subunit RPB9.

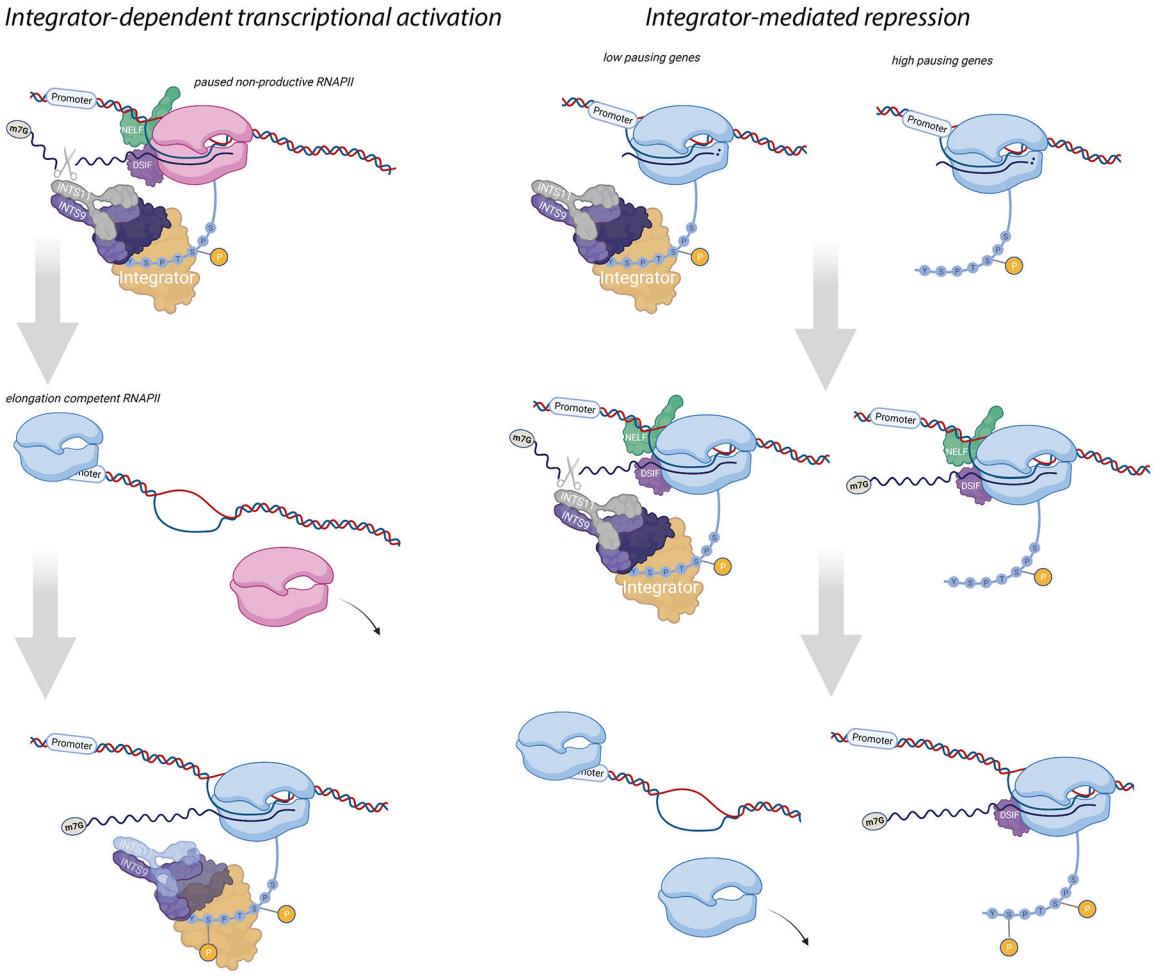

Figure 4. Cleavage of nascent pre-mRNA at RNA polymerase II pausing sites.

Integrator regulates the release and stability of RNA polymerase II (Pol II) pausing at promoter-proximal sites through its cleavage of the nascent mRNA. Conflicting models may explain the role of Integrator in transcriptional gene regulation.

A) Cleavage of nascent RNAs has been proposed to elicit gene activation by clearing away long-paused Pol II complexes that are not competent for productive elongation. Elongation-competent Pol II holoenzymes can be effectively recruited only after cleavage of the nascent pre-RNA by Integrator and premature termination of the lingering polymerase. Elongation-competent Pol II are successfully licensed for productive transcript elongation past the pausing site, without any further Integrator-mediated cleavage occurring.

B) Alternatively, Integrator has been proposed to preferentially target a subset of genes, enforcing post-initiation transcriptional repression. Genes recruiting high levels of Integrator maintain lower levels of promoter-proximal Pol II by continuously dislodging paused polymerase and therefore curtailing productive elongation (left). Conversely, genes recruiting less Integrator maintain stable levels of paused Pol II and can support productive elongation (right). In all models, the cleavage activity of Integrator at the pausing site of protein-coding genes is believed to occur following capping (m^7G) of the nascent pre-mRNA.

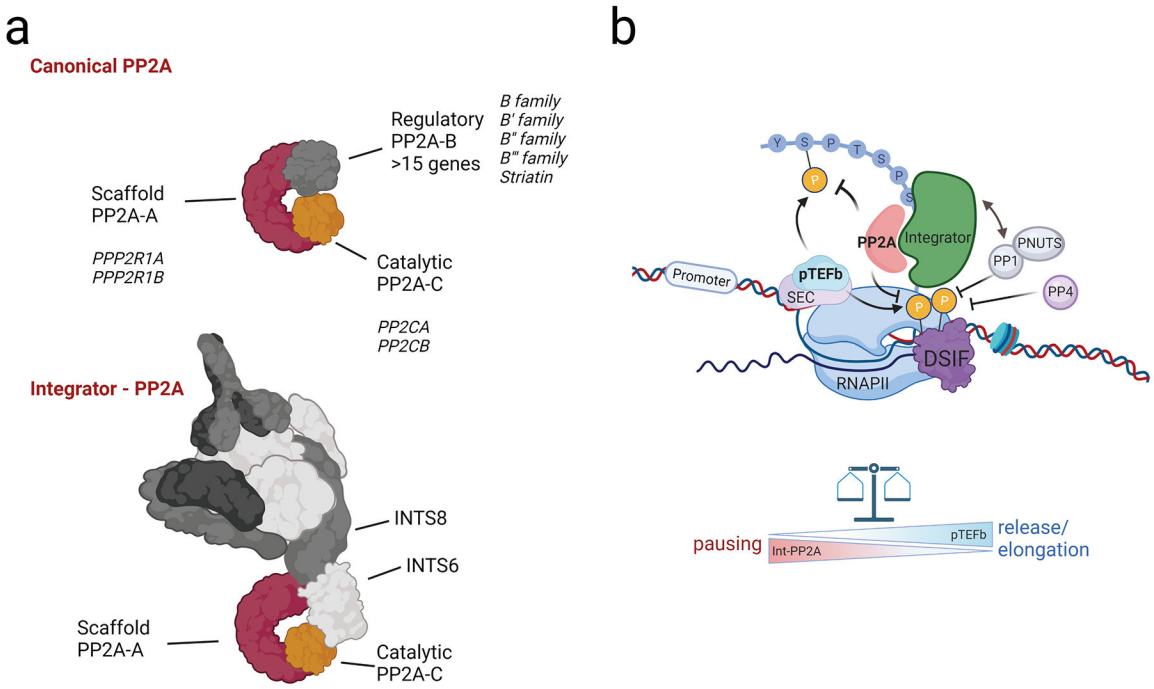

Figure 5. Integrator with the phosphatase module inhibits pause-release of RNA polymerase II.

A) Canonical serine/threonine-protein phosphatase 2A (PP2A) complexes are composed of PP2A-A (scaffold subunit), PP2A-C (catalytic subunit) and PP2A-B (regulatory subunit) (Top). Although PP2A-A–PP2A-C heterodimers are active in vitro, their efficient targeting in vivo requires assembly with PP2A-B. The Integrator complex is frequently found associated with PP2A (Int–PP2A), together also known as INTAC. In the Int–PP2A complex, PP2A-B is replaced by INTS6, while INTS8 contacts the N-terminus of PP2A-A (bottom). All other known Integrator subunits are stably associated with Int–PP2A.

B) The phosphatase module of Integrator inhibits promoter-proximal pause-release of RNA polymerase II (Pol II) by antagonizing phosphorylation of several targets of the kinase positive transcription elongation factor b (pTEFb; cyclin-dependent kinase 9–cyclin T1). Recruitment of pTEFb to Pol II is thought to enable pause-release and productive transcript elongation at protein-coding genes. Int–PP2A and pTEFb compete for a number of target residues, including at the Pol II C-terminal domain (CTD; most notably Ser2), and Ser666 and Thr806 of SPT5 (subunit of DRB-sensitivity inducing factor (DSIF)). In addition to Int–PP2A, the phosphatases PP4 and PP1–PNUTS also dephosphorylates SPT5. Notably, the PP1 phosphatase was found biochemically to be associated with Integrator.

PNUTS, serine/threonine-protein phosphatase 1 regulatory subunit 10; SEC, super elongation complex.

Table 1.

The functions of Integrator at different non-coding RNAs

RNA SPECIES	FUNCTION of INTEGRATOR	CELL TYPE, ORGANISM	REFS
Short RNAs (<0.2kb)			
U snRNAs	3'-box-mediated cleavage	All cell types, various metazoans	21,23,25,59,63
Viral miRNAs	Transcript release and maturation	Infected lymphocytes, marmoset	83,84
piRNAs	3' cleavage (unknown motif)	Germ cells, *Caenorhabditis elegans*	87,88
Long RNAs (>0.2kb)			
eRNAs	3' cleavage (at unknown motif)	Multiple cell types, *human*	58,81,82
TERC	3' cleavage (at unknown motif)	Human cell lines	92
NEAT1	Supporting early transcription termination	Human cell lines	94
Other lncRNAs	Supporting early transcription termination	Human cell lines	82

eRNA, enhancer RNAs; lncRNAs, long non-coding RNAs; miRNAs, microRNAs; NEAT1, nuclear paraspeckle assembly transcript 1; piRNAs, PIWI-interacting RNAs; snRNAs, small nuclear RNAs; TERC, telomerase RNA template component.

Table 2.

Integrator mutations in human development

SUBUNIT	MOLECULAR FUNCTION	MAIN PHENOTYPE	REFS
INTS1 (biallelic)	Central backbone, provides scaffolding to catalytic modules	Cognitive delay, absence of speech, cataracts and/or glaucoma, facial dysmorphism	150,151
INTS8 (biallelic)	Phosphatase module, required for PP2A recruitment	Cognitive delay, absence of speech, motor impairment	140
INTS13	Enhancer module, not required for cleavage or phosphatase activity	Oral-facial-digital anomalies, speech abnormality	132

PP2A, serine/threonine-protein phosphatase 2A