ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

BIOLOGIA CELLULARE E MOLECOLARE

Ciclo 37

**Settore Concorsuale:** 05/I1 - GENETICA

**Settore Scientifico Disciplinare:** BIO/18 - GENETICA

# THE ROLE OF INTS3 IN TRANSCRIPTIONAL REGULATION

**Presentata da:** Ilan Kirkel

**Coordinatore Dottorato**

Anna Maria Porcelli

**Supervisore**

Giovanni Perini

**Co-supervisore**

Alessandro Gardini

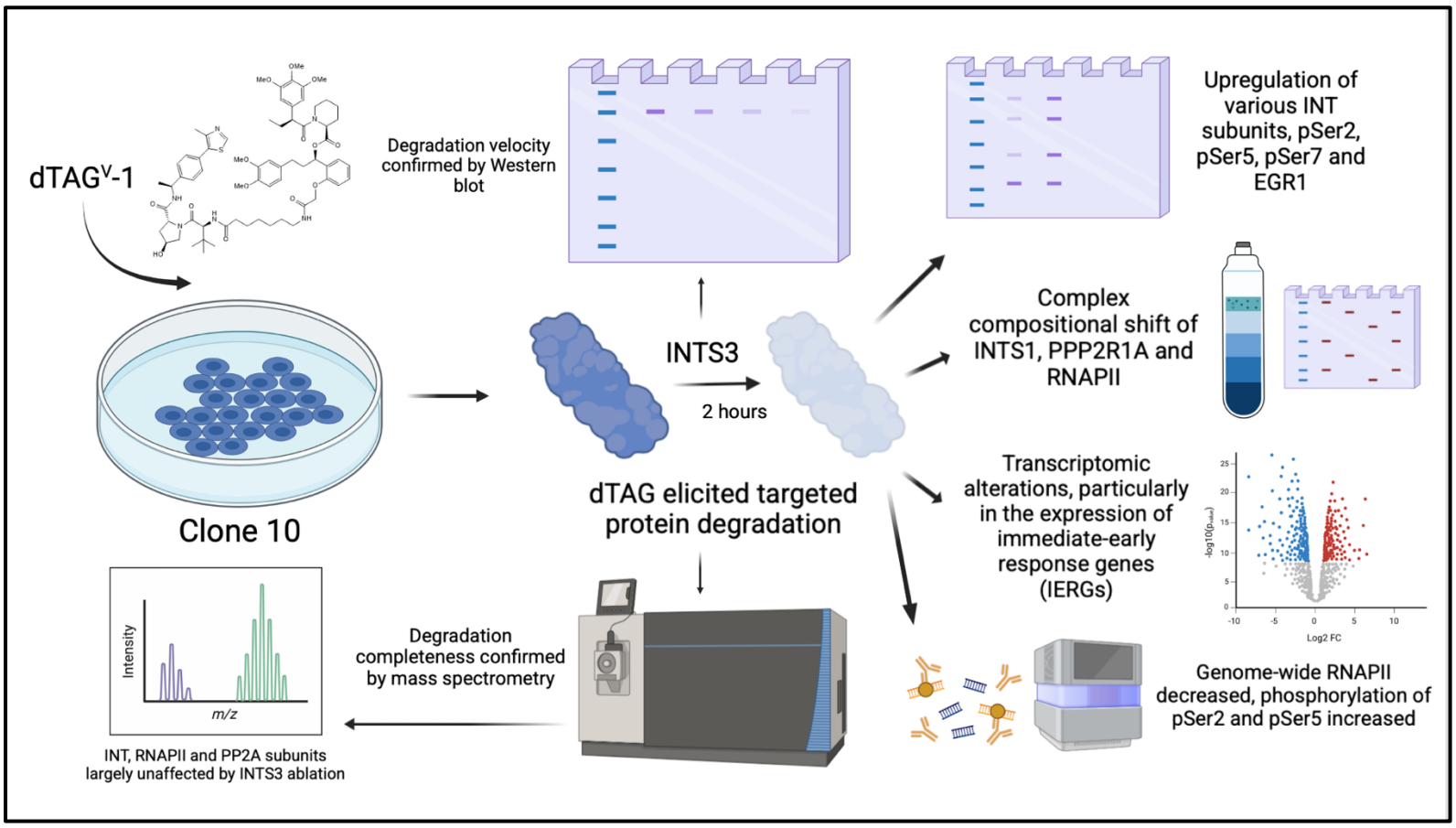
Esame finale anno 2025

# *Abstract*

Integrator (INT) is an RNA polymerase II-associated transcription complex, comprising fifteen subunits clustered into distinct functional modules. Each module individually contributes to Integrator's various regulatory roles in transcription which include modulation of RNA Pol II (RNAPII) pause-release transition, endonucleolytic cleavage and processing of nascent RNA species, targeted dephosphorylation of basal transcriptional machinery, and DNA repair. In particular, Integrator subunit 3 (INTS3) has been implicated in DNA damage repair in conjunction with the Sensor of ssDNA (SOSS) complex, alongside a role in preventing reassociation of RNAPII following premature transcriptional termination.

However, a satisfactory model of the mechanism by which INTS3 contributes to overall transcriptional regulation still remains elusive. In this light, a dTAG targeted protein degradation (TPD) system for INTS3 was generated in OVCAR8 cells to probe the rapid effects of INTS3 ablation on transcriptional regulation. The degradation of INTS3 is observed to upregulate the expression of various INT subunits and phosphoserine states of Ser2, Ser5 and Ser7 of the carboxyterminal domain (CTD) of RNAPII, and additionally may have an effect on INTAC (INT conjugated with PP2A) complex formation. Furthermore, INTS3 ablation is correlated with changes to the transcriptome with respect to the upregulation of immediate early response genes (IERGs). In a wider context, SOSS complex constituents were found to colocalize with INTS3 across the genome, predominantly at the AP-1 promoter element which is responsible for IERG expression. Moreover, it was observed that splicing factor 3 (SF3) subunits are differentially pulled down dependent on INTS3 ablation. Lastly, INTS3 degradation results in the reduction of RNAPII across the genome alongside an increase in pSer2 and pSer5 states, as well as a more efficient release of RNAPII into the gene body indicated by a decreased traveling ratio for RNAPII. Overall, these results implicate INTS3 in a variety of transcriptional regulatory events from affecting the INT complex directly, to changing RNAPII processivity across the genome, to broader changes to the transcriptome.

Graphical Abstract



# *Index*

*Abstract* ..... *1*
Graphical Abstract ..... 2
*Index* ..... *3*
List of Figures ..... 5
*Chapter 1: Introduction* ..... *7*
Transcriptional regulation ..... 7
The Integrator Complex ..... 11
The SOSS Complex ..... 16
Integrator Subunit 3 (INTS3) ..... 17
Targeted Protein Degradation Systems ..... 18
Research Aims ..... 23
*Chapter 2: Results* ..... *24*
An INTS3 dTAG system was generated for rapid and reliable ablation of INTS3 at the protein level 24
Integrator subunits and module expression, and CTD phosphoserine state, is affected by INTS3 ablation ..... 36
INTS3 ablation is correlated with changes to the transcriptome, especially with respect to immediate-early response genes (IERGs) ..... 42
Splicing factor 3 subunits are differentially pulled down dependent on INTS3 ablation ..... 51
The SOSS complex is present alongside INTS3 across the genome, found predominantly at a motif which controls IERG transcription ..... 53
INTS3 ablation affects RNAPII occupancy and increases pSer2 and pSer5 states across the genome. 58
*Chapter 3: Discussion* ..... *64*
Rigor and Reproducibility ..... 69
Future Directions ..... 71
*Chapter 4: Methods* ..... *75*
Reagents ..... 75
Cell lines ..... 78
Generation of CRISPR-Cas9 sgRNA and homology directed repair (HDR) plasmids ..... 78
*CRISPR-Cas9 sgRNA plasmids* ..... 78
*Homology directed repair (HDR) plasmids* ..... 80
Transfection of OVCAR8 cells and colony picking ..... 81

SDS-PAGE and immunoblotting ..... 81
Co-immunoprecipitation ..... 82
Nuclear extraction ..... 83
Glycerol gradient ..... 84
IP followed by mass spectrometry ..... 84
Quant-seq ..... 85
Gene ontology analysis ..... 86
Chromatin immunoprecipitation followed by sequencing (ChIP-seq) ..... 86
*References* ..... *89*

# List of Figures

Figure 1: The CTD phosphorylation code throughout the transcription cycle ..... 7

Figure 2: RNAPII recruitment to the TSS, promoter proximal pausing, and release to productive elongation..... 9

Figure 3: Integrator subunits and known interacting partners ..... 13

Figure 4: Structure of the SOSS complex..... 17

Figure 5: INTS3 as a constituent of both Integrator and the SOSS complex ..... 18

Figure 6: The auxin inducible degron (AID) system..... 20

Figure 7: The degradation TAG (dTAG) system..... 22

Figure 8: Schematic of insert for HDR plasmids and synonymous mutations present in LHA ... 25

Figure 9: Schematic of gRNA sequences to cleavage sites along INTS3..... 26

Figure 10: HDR plasmids constructed for insertion of FKBP12F36V by homology directed repair ..... 28

Figure 11: CRISPR/Cas9 knock-in validation by PCR if gDNA from OVCAR8 cells ..... 30

Figure 12: Immunoblots for dTAG induced degradation for clones C1 and C4..... 31

Figure 13: Western blot for long- and short-time course of INTS3 degradation by dTAG treatment ..... 33

Figure 14: Short-time course of INTS3 degradation after treatment with dTAG for clones 3, 16, and 10..... 34

Figure 15: Co-immunoprecipitation pull down of HA-tag for clone 10..... 36

Figure 16: Integrator, phosphoserines of the CTD of RNA Pol II, PP2A, NABP1 and EGR1 change in signal detected by Western blot post-dTAG treatment for four hours..... 37

Figure 17: Integrator, phosphoserines of the CTD of RNA Pol II, PP2A, NABP1 and EGR1 change in signal detected by Western blot post-dTAG treatment for eight hours ..... 38

Figure 18: Glycerol gradient of nuclear extract for clone 10 cells untreated and after four hours dTAG treatment ..... 41

Figure 19: PCA plot of clone 10 cells treated with dTAG ..... 43

Figure 20: Volcano plots differentially expressed genes for 30 minutes and 1 hour dTAG treatment vs. untreated ..... 46

Figure 21: PCA plot of wtOVCAR8 cells treated with dTAG ..... 47

Figure 22: Volcano plots differentially expressed genes for 30 minutes and 1 hour dTAG treatment vs. untreated ..... 49

Figure 23: Gene ontology analysis for upregulated gene for 30 minutes and one-hour dTAG treated cells vs. untreated ..... 50

Figure 24: Immunoprecipitation followed by mass spectrometry for untreated against four hours dTAG treated samples ..... 52

Figure 25: Representative gene tracks for RNA Pol II, INTS3 and NABP1 binding across the genome ..... 54

Figure 26: Average profile of INTS3 and NABP1 binding across active genes in OVCAR8 ..... 55

Figure 27: Promoter and enhancer binding proportions for INTS3 and NABP1 ..... 56

Figure 28: De novo and known motif discovery for intersecting motifs of INTS3 and NABP1 . 57

Figure 29: Average profile of RNAPII occupancy across the transcription start site (TSS), gene body, and transcription end site (TES) in clone 10 cells ..... 60

Figure 30: Average profile of pSer2 phosphorylation across the transcription start site (TSS), gene body, and transcription end site (TES) in clone 10 cells ..... 61

Figure 31: Average profile of pSer5 phosphorylation across the transcription start site (TSS), gene body, and transcription end site (TES) in clone 10 cells ..... 62

Figure 32: Traveling ratio of RNAPII after dTAG treatment of 0, 30 and 60 minutes in clone 10 cells ..... 63

# *Chapter 1: Introduction*

## Transcriptional regulation

Transcriptional regulation is fundamental to the homeostasis and proliferation of the cell. Although each cell contains an entire genome, only a restricted subset of genes is expressed at any instance. Such cell-specific spatiotemporal expression ultimately confers individual cellular identity, function, and developmental trajectory. In this light, multicellular eukaryotes depend on tightly regulated transcription to give rise to the multitude of discrete cell types required for the optimal function and development of the entire organism. Within the nucleus, structurally and functionally conserved molecular machinery ensures fidelitous transcription of DNA to mature mRNA – a process orchestrated through additional specialized regulatory networks. In particular, the twelve-subunit holoenzyme RNA polymerase II (RNAPII) functions as the core complex involved in the transcription of protein-coding genes alongside a large portion of snRNAs (Corden, 2013; Moreno *et al.,* 2023). RPB1, the largest subunit of the complex, contains the DNA binding domain and forms the core for RNAPII catalytic activity (Jeronimo *et al.,* 2016). Notably, the disordered carboxy-terminal domain (CTD) of RPB1 comprises a highly conserved heptad sequence ($Y_1S_2P_3T_4S_5P_6S_7$) of 26 – 52 repeats which undergoes post-translational modification dependent on the stage of transcription (Zhang and Corden, 1991; Moreno *et al.,* 2023) (Figure 1). In association with a host of protein complexes, RNAPII is guided through the three serial stages of transcription, namely: initiation, elongation and termination.

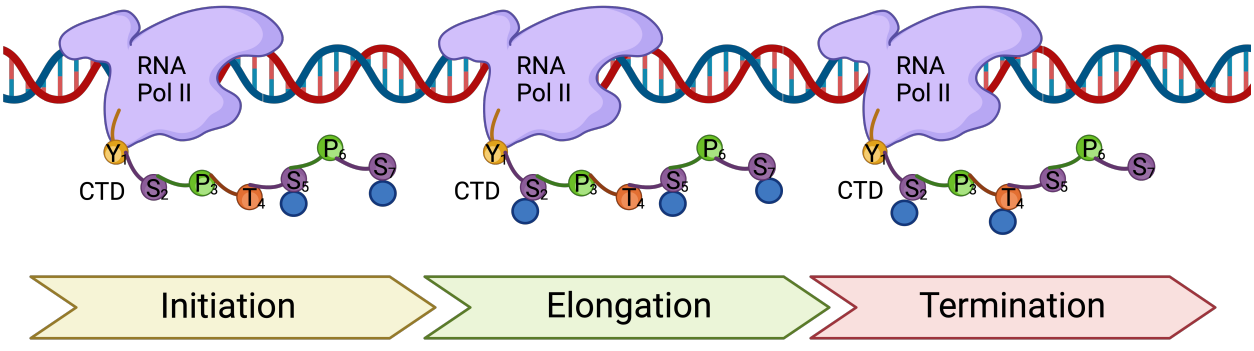



Figure 1: The CTD phosphorylation code throughout the transcription cycle. At the start of transcriptional initiation, the CTD of RNAPII is phosphorylated (blue circles) at Ser5 and Ser7.

Phosphorylation of Ser2 switches RNAPII into an actively elongating complex. As RNAPII elongates through the gene body, the phosphate groups on Ser5 and Ser7 are gradually removed by phosphatases. As RNAPII approaches the poly(A) site, termination factors are recruited, Thr4 is phosphorylated, and one transcription cycle is complete. Created in BioRender.

At the start of transcription, RNAPII is assembled into a pre-initiation complex (PIC) comprising RNAPII, the co-activator Mediator complex (consisting of 26 subunits), and six multi-subunit general transcription factors (gTFs) (Flanagan *et al.*, 1991; Schier and Taatjes, 2020; Richter *et al.*, 2022). Transcription is initiated through RNAPII recognition of promoter DNA from the vast excess of non-promoter DNA across the genome (Young *et al.*, 2002). Initiation proceeds through separation of the DNA duplex to form a "transcription bubble" thereby exposing the template strand (Young *et al.*, 2002; Zhang *et al.*, 2012). RNAPII binding occurs upstream of the gene body alongside association with gTFs (Adelman and Lis, 2012) (Figure 2). Subsequently, RNAPII enters the early stage of initial elongation, passing the transcription start site (TSS) and accumulates at elevated levels in the promoter-proximal region, 30 – 60 nucleotides downstream of the TSS in a paused state (Jonkers and Lis, 2015) (Figure 2). Such promoter-proximal pausing acts as a crucial rate-limiting step for transcription, which is subject to additional regulatory control, and operates as a quality checkpoint for transcript 5`-capping and RNAPII modification before productive elongation (Jonkers and Lis, 2015). Pausing of RNAPII at the promoter-proximal region is dependent on core promoter features that recruit RNAPII. These include various TFs that recruit DRB-sensitivity inducing factor (DSIF) and negative elongation factor (NELF) to stabilize paused RNAPII (Kwak and Lis, 2013), which themselves sterically exclude positive elongation factors such as the PAF complex and SPT6. Additionally, adjacent nucleosomes may contribute to pausing (Figure 2) (Kwak *et al.*, 2013).

The eventual release of paused RNAPII is facilitated by the positive transcription elongation factor (P-TEFb) complex, consisting of cyclin T1 and cyclin-dependent kinase 9 (CDK9) (Zhou *et al.*, 2012), whereby P-TEFb acts as a key regulator of early elongation (Adelman and Lis, 2012) (Figure 2). Specific cofactors and TFs recruit P-TEFb to the promoter region occupied by RNAPII, NELF and DSIF, resulting in the phosphorylation of NELF (leading to its eviction from the complex), Ser2 of the CTD, and DSIF, which is maintained in the complex as a positive elongation factor (through the triggering of allosteric changes) (Kwak and Lis, 2013; Kwak *et al.*, 2013; Welsh and Gardini, 2023). In this light, P-TEFb exists in one of two states: as a component of an

inhibitory complex, or as an active complex which phosphorylates the RNAPII CTD and pausing factors (Zhou *et al.*, 2012). Therefore, the degree of RNAPII pausing is contingent upon the equilibrium between pausing factors (such as DSIF, NELF, and the +1 nucleosome in conjunction with core promoter elements) and activating factors, which either facilitate the recruitment of P-TEFb to paused RNAPII or directly activate P-TEFb.

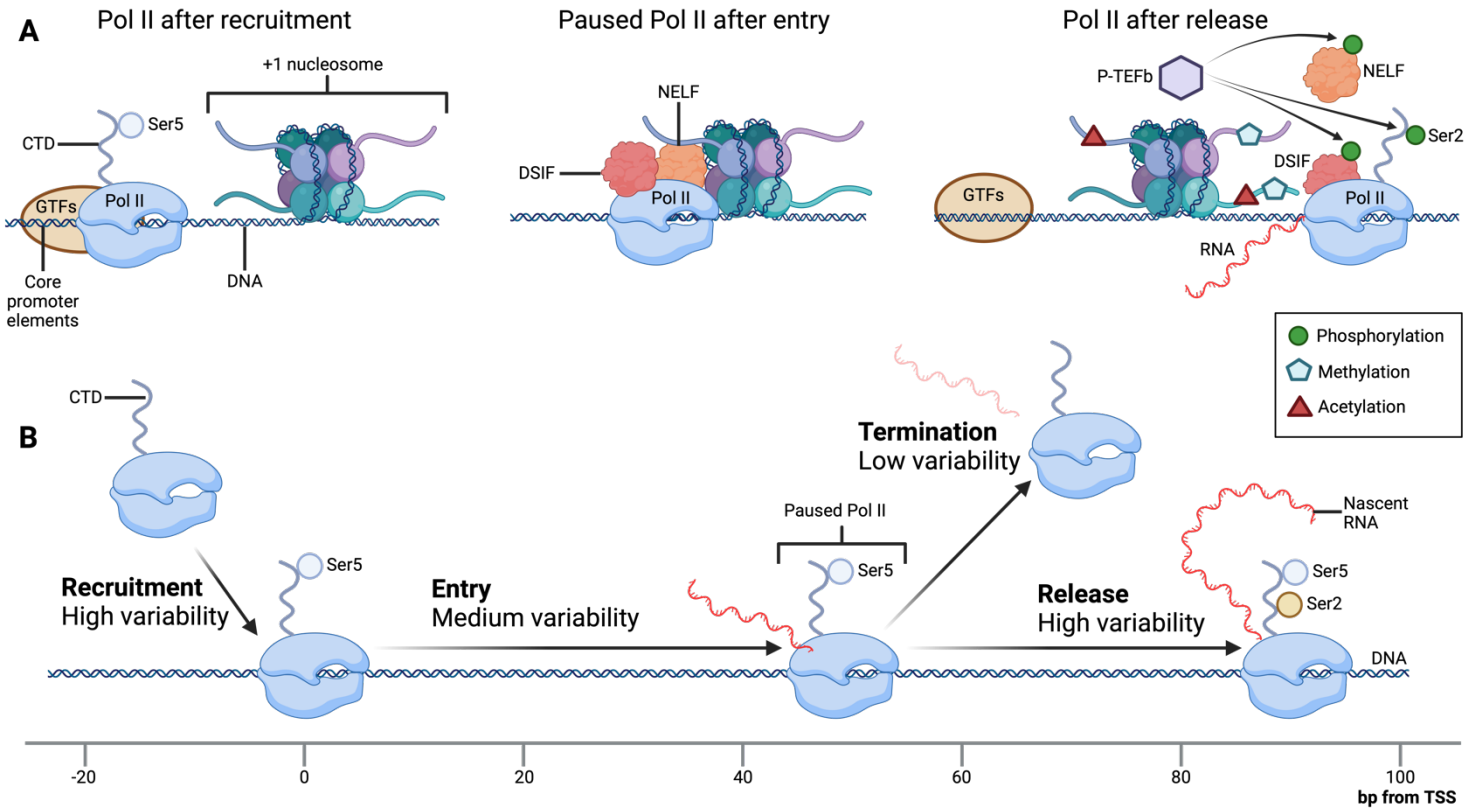



Figure 2: RNAPII recruitment to the TSS, promoter proximal pausing, and release to productive elongation. (A) RNAPII associates with transcription factors at the promoter element downstream of the TSS. Depending on factors that underlie RNAPII recruitment, initiation, pausing and release, the position of RNAPII along the promoter and the composition of the complex will vary. The pre-initiation complex (PIC) results from the recruitment of general transcription factors (gTFs). After swift RNAPII initiation and entrance into the pause site, core promoter elements, the +1 nucleosome, alongside currently negative elongation factors DSIF and NELF, facilitate pausing of the polymerase. The release of paused RNAPII is mediated by P-TEFb phosphorylation of NELF, DSIF, and the CTD of RNAPII, whereby DSIF transitions to a positive elongation factor following phosphorylation. (B) Transcription is primarily regulated near the TSS, during the recruitment of RNAPII to promoters, and at the release of RNAPII from the promoter proximal pause site. There is variability in terms of rate for each of these steps. Other steps of transcription are not as variable in terms of their rate of incidence such as transcriptional initiation and pause site entry, as well as the premature termination of transcription from the pause site (Jonkers and Lis, 2015). Created in BioRender.

Productive elongation commences following RNAPII release from the promoter-proximal pause site, thus further elongating the nascent RNA transcript (Veloso *et al.*, 2014). Rates of elongation differs between and within genes – up to threefold – with these rates seeming to affect co-transcriptional processes such as the maintenance of genome stability, splicing, and transcriptional termination (Veloso *et al.*, 2014). Various additional factors such as histone marks and gene features, such as exon number, can further modulate elongations rates (Fuchs *et al.*, 2014). Productive elongation is less efficient within the first kilobases of mammalian genes, evident by a rate increase from 500 bases per minute within the first few kilobases to 2 - 5 kilobases per minute after approximately 15 kilobases (Jonkers *et al.*, 2014; Veloso *et al.*, 2014). This may be a result of the progressive accumulation and modification the transcriptional machinery as elongation progresses (Jonkers and Lis, 2015). For instance, maximum phosphorylation of Ser2 is not observed until a few kilobases into the gene body (Heidemann *et al.*, 2013). Furthermore, pausing factors such as DSIF, NELF and GDOWN1 are either gradually modified or evicted from the transcriptional machinery during elongation further affecting elongation rate. This incremental "maturation" of the transcription machinery may allow for the recruitment of accessory factors crucial for co-transcriptional events such as RNA splicing (Jonkers and Lis, 2015). Notably, the paused-to-productive transcription state is not a simple switch: even after extensive transcription of the gene body, further hinderances may continue to affect elongation rates. These impediments include the presence of exons, and mRNA cleavage and polyadenylation sites (Gromak *et al.*, 2006; Kwak *et al.*, 2013). Thus, the dynamic changes in elongation rate regulates transcription throughout its travelling along the gene body.

Once RNAPII has transcribed beyond the transcription end site (TES), transcriptional termination occurs whereby pre-mRNA is cleaved and processed. Upon final processing, the mature mRNA translocates to the cytoplasm where eventual translation results in functional protein. Notably, our current understanding of eukaryotic transcription is largely founded on studies in unicellular *Saccharomyces cerevisiae* which may restrict our accurate understanding of transcription in higher, multicellular eukaryotes. Indeed, it appears multicellular eukaryotes rely on more complex transcriptional regulatory mechanisms as a result of the need for differentiation into multiple, distinct cell types. One such instance of "higher" transcriptional machinery – specifically a

transcriptional co-regulator absent in yeast and rather confined to higher eukaryotes – is the Integrator complex.

## The Integrator Complex

The Integrator complex is a recently identified multi-subunit complex of at least 15 known subunits within mammalian cells, capable of binding the CTD of RNAPII (Baillat *et al.*, 2005) (Figure 3). Integrator is only observed in metazoans and is absent in yeast, unlike most other known transcriptional co-regulator complexes, intimating a more recent evolution of the complex (Welsh and Gardini, 2023). The discovery of Integrator was a result of identification, by mass spectrometry, of twelve then-uncharacterized open reading frames (ORFs) which associated with multiple RNAPII subunits (Baillat *et al.*, 2005; Welsh and Gardini, 2023). Orthologues of all twelve subunits were identified across *metazoa* (but absent from yeast), suggesting the exclusivity of the complex to multicellular eukaryotes (Baillat *et al.*, 2005). Moreover, orthologues of Integrator subunits 11 and 9 (INTS11 and INTS9) have been observed in plants (Liu *et al.*, 2016), but absent in fungi, further supporting the hypothesis that Integrator is central to a transcription regulation network that favours multicellularity (Welsh and Gardini, 2023).

Initially, functionality of the Integrator complex was proposed by sequence homology-based annotation tools: a β-lactamase-β-CASP domain present within Integrator subunits INTS11 and INTS9 showed high homology with cleavage and polyadenylation factor subunits 100 and 73 (CPSF100 and CPSF73), respectively, hinting towards RNA endonuclease activity for either of these subunits (Dominski *et al.*, 2005; Welsh and Gardini, 2023). Indeed, knockdown of the largest scaffolding subunit of Integrator, INTS1, or of the presumed (although now confirmed) catalytic subunit INTS11, resulted in the over-accumulation of unprocessed uridylate-rich small nuclear RNA (U snRNAs) thus forwarding the notion that Integrator is involved alongside RNAPII in the processing and cleavage of nascent transcripts (Baillat *et al.*, 2005; Albrecht and Wagner, 2012; Welsh and Gardini, 2023). Functional and structural analyses have confirmed that INTS11 and INTS9 form a pseudosymmetric heterodimer with outward facing catalytic clefts, whereby only the INTS11 is catalytically active as INTS9 lacks zinc-binding ability necessary for enzymatic activity (Pfleiderer and Galej, 2021). An additional subunit, INTS4 – a structural homologue of symplekin in the CPSF complex – stabilizes the interaction between INTS11 and INTS9 through

multiple points of interface (Albrecht *et al.*, 2018; Pfleiderer and Galej, 2021). This trimeric module (INTS11, INTS9 and INTS4) forms the catalytic core of Integrator and is termed the endonuclease module. Along with stabilizing the INTS11/9 heterodimer, INTS4 acts to anchor the endonuclease module to the remainder of the Integrator complex through two separate domains contacting INTS1 and INTS7 – two subunits comprising the backbone of the complex (Zheng *et al.*, 2020; Fianu *et al.*, 2021). The Integrator complex itself binds paused RNAPII by making widespread contacts with the RNAPII core complex as well as the CTD (Zheng *et al.*, 2020). When Integrator is bound, the endonuclease module is situated adjacent to the RNA exit site of RNAPII and acts to cleave nascent RNA approximately 20-bp from the RNAPII active site, thus contributing to nascent RNA processing or transcriptional attenuation (Sabath *et al.*, 2020; Zheng *et al.*, 2020; Welsh and Gardini, 2023). Apart from INTS11/9/4 of the endonuclease module, various other Integrator subunits group into modules which confer additional functionality to the complex, namely: the enhancer (or 'arm') module, and the phosphatase module.

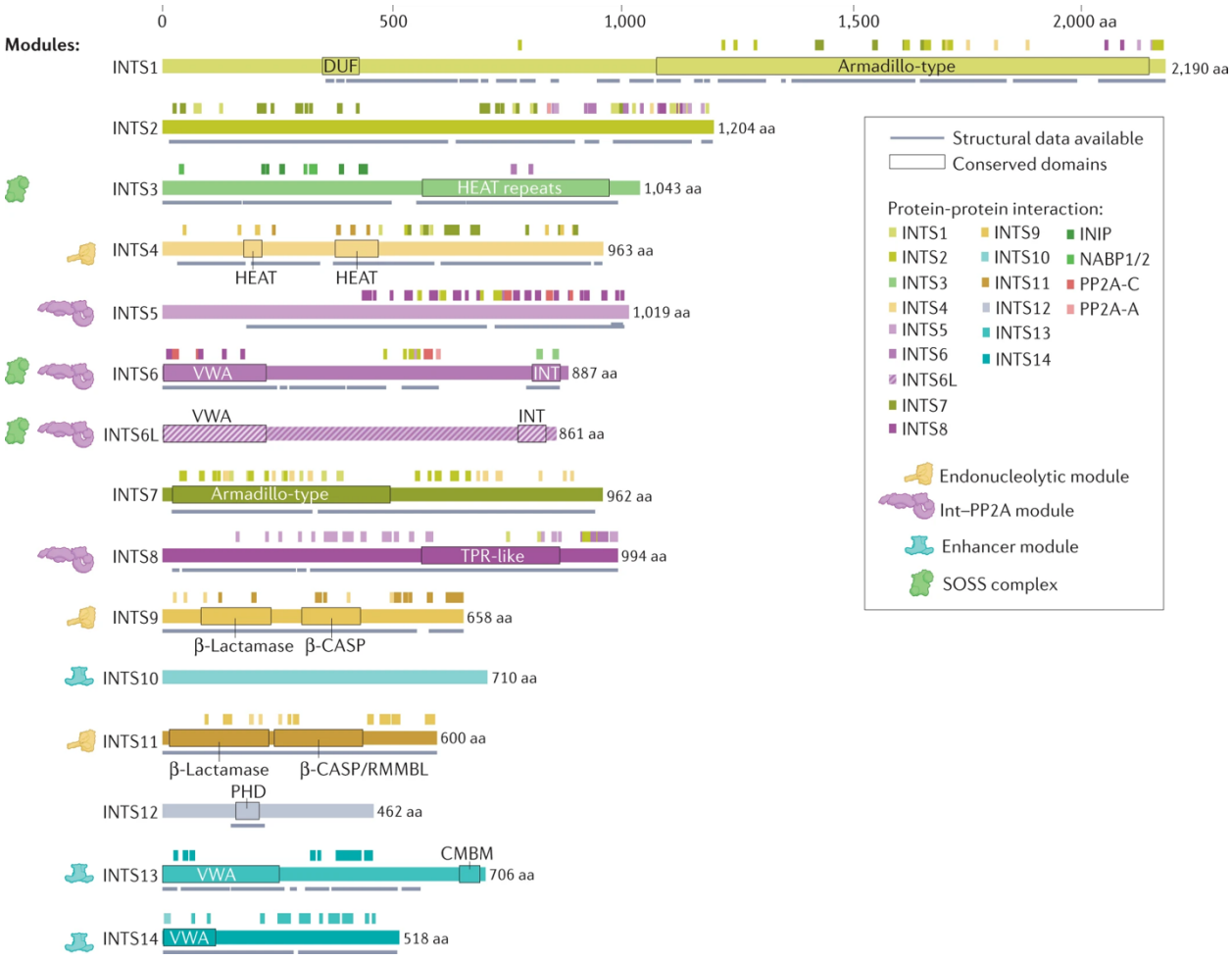



Figure 3: Integrator subunits and known interacting partners. All 15 subunits of the Integrator complex are shown to scale. Annotated protein motif and domains are titled and boxed. Subunits are grouped according to module they are located: the endonucleolytic module (yellow), phosphatase module (purple), enhancer module (blue) or SOSS complex (green). Abbreviations are as follows: aa – amino acid; CMBM – cleavage module binding motif; DUF – domain of unknown function; VWA – von Willebrand factor type A domain. (Welsh and Gardini, 2023).

Integrator subunits 10, 13 and 14 (INTS10, INTS13 and INTS14) form a distinct functional module within the Integrator complex, termed the enhancer module (Sabath *et al.*, 2020) (Figure 3). Initially, INTS13 and INTS14 were not recovered alongside the twelve first identified subunits of the Integrator complex; instead, they were tentatively shown to be part of Integrator following functional screening in *D. melanogaster* (Chen *et al.*, 2012; Sabath *et al.,* 2020; Welsh and Gardini, 2023). INTS10, INTS13 and INTS14 comprise a trimeric module that associates with the complete Integrator-RNAPII complex, whereby INTS13 and INTS14 are interlinked in a heterodimeric conformation and INTS10 binds the heterodimer at the vWA-domain of INTS14

(Sabath *et al.*, 2020; Pfleiderer *et al.*, 2021). The tightly entwined INTS13/14 heterodimer is unexpectedly similar to the Ku70-Ku80 DNA repair complex, which binds DNA double stranded break ends, suggesting an evolutionary and functional role in nucleic acid processing for INTS13/14 (Sabath *et al.*, 2020). Functionally, the INTS10-INTS13-INTS14 module has a supplementary function in snRNA maturation alongside a more significant role on transcription termination (Sabath *et al.*, 2020). This is achieved through INTS13 interacting directly with the endonuclease module of the Integrator complex through its conserved C-terminal motif (Sabath *et al.*, 2020). This interaction is essential for snRNA processing, specifically, the 3'-end of spliceosomal small nuclear RNAs (snRNAs) (Sabath *et al.*, 2020). The binding of the INTS10/13/14 module to the endonuclease module likely brings the endonuclease module close to target transcripts, facilitating efficient transcript processing (Sabath *et al.*, 2020). In addition to snRNA processing and transcriptional termination, the enhancer module functions to determine cell fate. INTS13 functions as a monocytic/macrophagic differentiation factor (Barbieri *et al.*, 2018); through the targeting of early growth response 1 and 2 (EGR1/2) enhancers and their co-factor NAB2, INTS13 drives monocytic differentiation from progenitor cells (Barbieri *et al.*, 2018). In addition, depletion of INTS10 prevents proper neural differentiation in induced pluripotent stem cells (iPSCs) to neural progenitor cells and ultimately neurons (data yet unpublished). Overall, the enhancer module confers a crucial multifaceted role to the Integrator complex by virtue of RNA processing, transcriptional regulation and cell fate determination.

In addition to the aforementioned functions of Integrator, the complex further regulates RNAPII transcription in association with the cellularly ubiquitous phosphatase: protein phosphatase 2A (PP2A), which comprises part of the phosphatase module of the Integrator complex (Lambrecht *et al.*, 2013; Cossa *et al.*, 2021; Welsh and Gardini, 2023) (Figure 3). PP2A is a serine/threonine phosphatase which is highly conserved across all cell types and contributes to the greater proportion of phosphatase activity in any given cell (Lambrecht *et al.*, 2013; Seshacharyulu *et al.*, 2013; Cossa *et al.,* 2021). PP2A is comprised as a heterotrimeric complex of a scaffolding subunit (PP2A-A), regulatory subunit (PP2A-B), and catalytic subunit (PP2A-C) (Lambrecht *et al.*, 2013). As a heterotrimer, PP2A specifically targets and dephosphorylates a vast spectrum of proteins, ranging from intracellular signaling mediators to mitotic spindle components (Seshacharyulu *et al.,* 2013). Initial evidence for association of PP2A with the Integrator complex arose from

proteomics studies implicating a robust interaction between subunits of both complexes (Malovannaya *et al.*, 2010; Malovannaya *et al.*, 2011). This was confirmed through cryo-electron microscopy which resolved nine subunits of the Integrator complex alongside a heterodimer of the PP2A complex (PP2A-A and PP2A-C; notably PP2A-B was absent), with both complexes together termed "INTAC" (Zheng *et al.*, 2020). The structure of INTAC showed the PP2A heterodimer anchored to Integrator primarily by Integrator subunits 8 and 6 (INTS8 and INTS6) which are most crucial for association of PP2A with Integrator (Zheng *et al.*, 2020; Fianu *et al.*, 2021; Vervoort *et al*., 2021). INTS6 tethers the PP2A heterodimer to the Integrator backbone while concurrently contacting NELF (Fianu *et al.*, 2021). Deletion of INTS6 results in loss of PP2A-A and PP2A-C from the Integrator complex while deletion of INTS8 leads to loss of both INTS6 and PP2A from the complex (Vervoort *et al*., 2021). This prevention of PP2A recruitment to the Integrator complex by depletion of INTS8 or INTS6 is generally correlated with an increase in transcription of nascent RNA and of steady-state RNA levels, indicating PP2A may play a role in regulating productive elongation (Huang *et al.*, 2020; Vervoort *et al*., 2021). In general, RNAPII pause-and-release activity downstream of the TSS is likely a result of the delicate equilibrium between kinases and phosphatases (such as PP2A) acting on the elongation complex (Vervoort *et al*., 2021). PP2A modulates elongation factors and phosphorylation of the RNAPII CTD throughout transcription which affects productive elongation and transcriptional pausing. For example, a loss of PP2A from INTAC facilitates increased phosphorylation of SPT5 (a constituent of DSIF) leading to an increase in productive elongation (Huang *et al.*, 2020; Vervoort *et al*., 2021). Phosphorylated serine residues of the RNAPII CTD (pSer2, pSer5 and pSer7) may act as additional substrates for PP2A as a part of INTAC, further indicating the role of INTAC in transcriptional pausing (Zheng *et al.*, 2020; Hu *et al.*, 2021; Vervoort *et al*., 2021). ChIP-seq data for PP2A-A and PP2A-C indicates their diffuse interaction with chromatin along the gene body and at the 3`-end of most genes, implying that PP2A regulates RNAPII activity beyond transcriptional pausing (Huang *et al.*, 2020; Vervoort *et al*., 2021). Altogether, the PP2A-Integrator interaction is critical for regulating RNAPII transcription by virtue of INTAC's broad role in modulating productive elongation and RNA processing.

In addition to PP2A, Integrator associates with the sensor of single-stranded DNA (SOSS) complex through contact with its third largest subunit: Integrator subunit 3 (INTS3), which will be discussed in later sections.

## The SOSS Complex

Various DNA repair proteins act to maintain the integrity and stability of the genome throughout DNA replication and RNA transcription. A particular risk to genome integrity is DNA double-stranded breaks (DSBs) – cytotoxic lesions which, if left unresolved or improperly repaired, may lead to genome instability and tumorigenesis (Bartek and Lukas, 2007; Huang *et al.*, 2009). Mammalian cells rely on two main DSBs repair programs: homologous repair (HR) and non-homologous end joining (NHEJ) (Kennedy and D'Andrea, 2006; Huang *et al.*, 2009). Depending on the nature of the DSB and the stage of the cell cycle, HR and NHEJ contribute differently to DSB repair (Sonoda *et al.*, 2006). The HR mechanism plays a crucial role in preserving genomic integrity by accurately repairing DNA double-strand breaks and facilitating the resumption of stalled or collapsed DNA replication forks. During the initial phases of HR, such DSBs undergo resection to produce single-stranded DNA (ssDNA). This ssDNA is then coated by single-stranded DNA binding proteins (SSBs), which are critical for various DNA processes including replication, recombination and repair in cells across all domains of life (Buis *et al.*, 2008; Hopkins and Paull, 2008). The major SSB in eukaryotes is the highly conserved replication protein A (RPA) which participates in critical processes such as DNA replication, transcription and repair (Huang *et al.,* 2009). However, there additionally exists two SSB homologues – NABP1 and NABP2 – which contribute to genome maintenance during various stages of the cell cycle. NABP1 is additionally aliased hSSB2, or SOSS-B1, and NABP2 is aliased hSSB1, or SOSS-B2. NABP2 plays a crucial role in the cellular response to DNA damage as cells deficient in the protein display errors in homologous recombination repair, defective G2/M checkpoint activation and increased sensitivity to ionizing radiation (Richard *et al.*, 2008). NABP1 and NABP2 are constituents of the heterotrimeric Sensor of ssDNA (SOSS) complex further comprising INTS3 (SOSS-A) and INIP (SOSS-C).

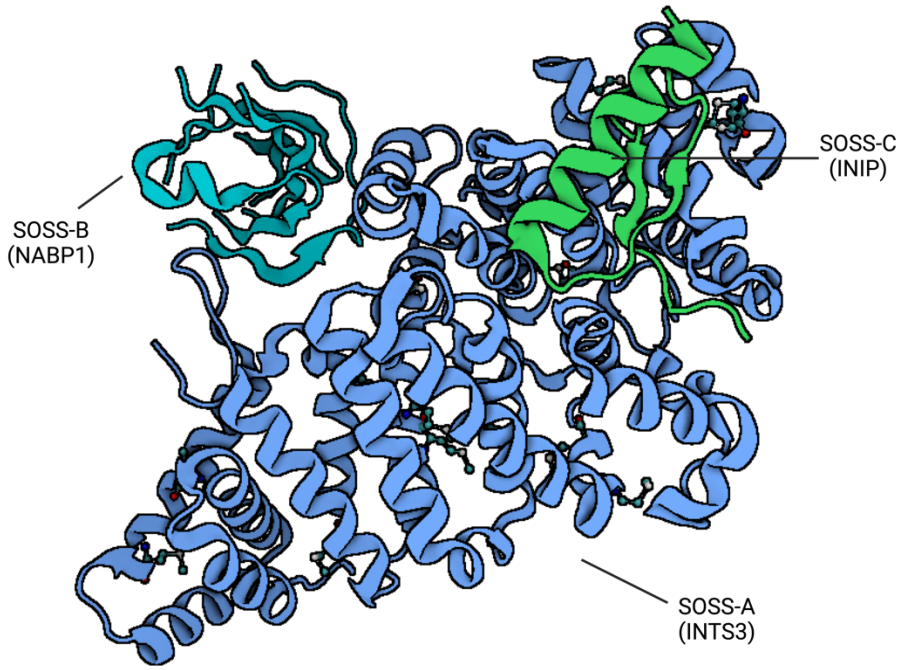



Figure 4: Structure of the SOSS complex. Schematic of the SOSS complex is shown: SOSS-A (INTS3) (sky blue), SOSS-B1 (NABP1) (turquoise), and SOSS-C (INIP) (green). SOSS-A acts as a scaffold for SOSS-B1 and SOSS-C to associate (PDB: 4OWT). Created with BioRender.

## Integrator Subunit 3 (INTS3)

INTS3 is the third largest subunit of the Integrator complex, with a molecular mass of 118 kDa (Welsh and Gardini, 2023). It is found as a constituent of both the Integrator complex and the SOSS complex. INTS3 is structurally comprised of two HEAT-repeat segments forming distinct N- and C-termini separated by an unstructured random coil linker (Li *et al.,* 2021). The C-terminus of INTS3 interacts with the C-terminus of INTS6 or INTS6L while the N-terminus interacts with the SOSS complex components (Figure 5). In a solved crystal structure and by chemical crosslinking, the C-terminus of INTS3 was observed to form a stable homodimer (Li *et al.,* 2021). A basic groove and cluster of conserved residues on opposing sides of the homodimer show binding affinity for ssDNA and ssRNA, and for association with INTS6 (Li *et al.,* 2021). INTS3 dimerization, however, is only necessary for nucleic acid binding as the protein is able to dock to INTS6 as a monomer (Li *et al.,* 2021). In particular, INTS3 binds at least 30-mer dT and shows similar affinity for random 30-mer ssDNA, however it does not bind dsRNA, dsDNA, or DNA/RNA hybrids (Li *et al.,* 2021). The α-helical structure of the N-terminal segment assembles NABP1, NABP2 and INIP (Ren *et al.,* 2014; Li *et al.,* 2021). As aforementioned, transient docking

of INTS3 to the Integrator complex prevents reassociation of RNA Pol II during premature transcriptional termination facilitated by Integrator (Fianu *et al.,* 2024). However, possible additional transcriptional effects of INTS3 remain elusive. In terms of cancer, *INTS3* is significantly overexpressed in hepatocellular carcinoma tumors and may play a role in the development or progression of hepatocellular carcinoma (Inagaki *et al.,* 2008). Furthermore, a recent CRISPR-Cas9 screening has identified INTS3 as an anti-apoptotic RNA binding protein and potential therapeutic target for colorectal cancer (Wang *et al.,* 2024).

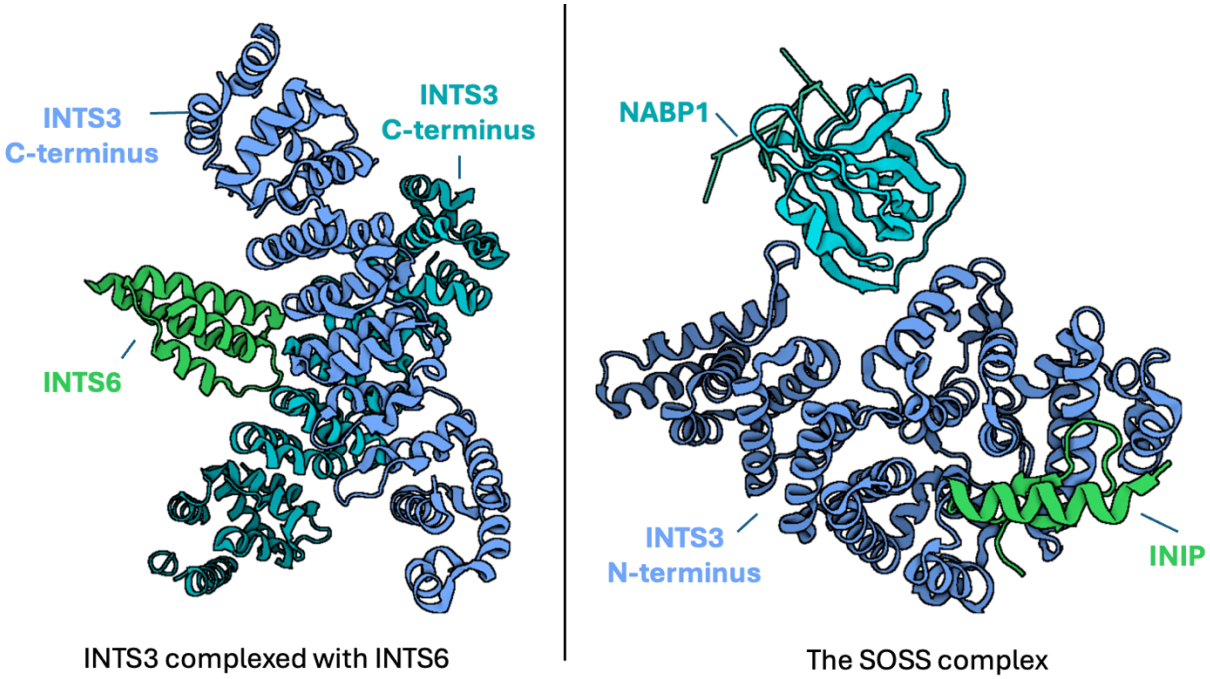



Figure 5: INTS3 as a constituent of both Integrator and the SOSS complex. The C-terminus of INTS3 forms a homodimer cleft (sky blue and turquoise) to allow docking with INTS6 (green) as a part of the Integrator complex. In addition, INTS3 (sky blue) acts as a scaffold for the SOSS complex comprising of both NABP1 (or NABP2) (turquoise) and INIP (green).

## Targeted Protein Degradation Systems

Gene perturbation strategies at the genomic or transcriptomic level, such as gene knockout by CRISPR/Cas9 or gene knockdown by RNA interference (RNAi), disrupt protein homeostasis and are robust techniques to dissect protein function in a cellular physiological context. However, each of these techniques are accompanied by significant drawbacks. Gene editing by CRISPR/Cas9, while leading to complete disruption of gene, leads to irreversible depletion of a protein of interest (POI) and cannot be used for essential genes; in comparison, RNAi can lead to incomplete

knockdown or a transient knockdown (Taylor *et al.,* 2017; Prozzillo *et al.,* 2019; Prozzillo *et al.,* 2020). Moreover, CRISPR/Cas9 and RNAi can give rise to off-target effects and does not affect all protein products in the cell, such as those already translated (Jackson *et al.,* 2003; Prozzillo *et al.,* 2020; Wang *et al.,* 2020). These strategies can additionally trigger compensatory mechanisms through their indirect depletion of a POI (Rossi *et al.,* 2015). Targeting a POI at the protein level potentially allows for the overcoming of these limitations: in particular, there is the advantage of acute degradation, alongside reversibility and a reduction of off-target events. This allows for the study of the primary effects of protein depletion removed from the secondary effects or adaptive responses triggered by the gradual silencing accompanying gene expression (Röth *et al.,* 2019). Several strategies for targeted protein degradation (TPD) have been developed to rapidly degrade a POI which exploit a degradation signal peptide sequence (tag) to appropriate E3 ubiquitin ligases to the POI leading to polyubiquitination and consequent proteasomal degradation by the ubiquitin-proteosome system (Prozzillo *et al.,* 2020). Two examples of such a system are the auxin inducible degron (AID) system and the dTAG degradation system.

Auxins such as IAA (indole-3-acetic acid) are phytohormones that control plant development and growth. As auxin levels increase, it permits transcription of auxin responsive genes through the degradation of transcriptional repressors AUX/IAA via the proteosome (Prozzillo *et al.,* 2020). Three essential components comprise this degradation pathway: the phytohormone, the target protein, and E3 ubiquitin ligase enzyme, itself being the SCF complex (SKP1, Cullin 1 and F-box) (Prozzillo *et al.,* 2020) (Figure 6). Depending on the application, the F-box subunit can be modified according to substrate specificity. One particular F-box – TIR1 – is able to bind auxins and lead to the recognition of the target (IAA/AUX). This recognition subsequently recruits an E2 ubiquitin-conjugating enzyme for polyubiquitination and degradation of the POI (Prozzillo *et al.,* 2020). The basis of the application of this TPD in non-plant cells is the presence of the SCF complex across all eukaryotes, although is limited by the missing TIR1 orthologues. In this light, TIR1 can be introduced exogenously to the cellular model followed by genetic manipulation of the target endogenous gene to incorporate an AID-tag sequence. The exogenously expressed TIR1 adaptor is able to conjugate with the SCF complex to form a functional SCF-TIR1 complex which can lead to degradation of the tagged POI by the proteosome in the presence of auxin (Prozzillo *et al.,*

2020). When treated with auxin, POI depletion is very rapid with a half-life of 10 – 20 minutes that may be reversed by auxin removal (Nishimura *et al.,* 2009; Holland *et al.,* 2012).

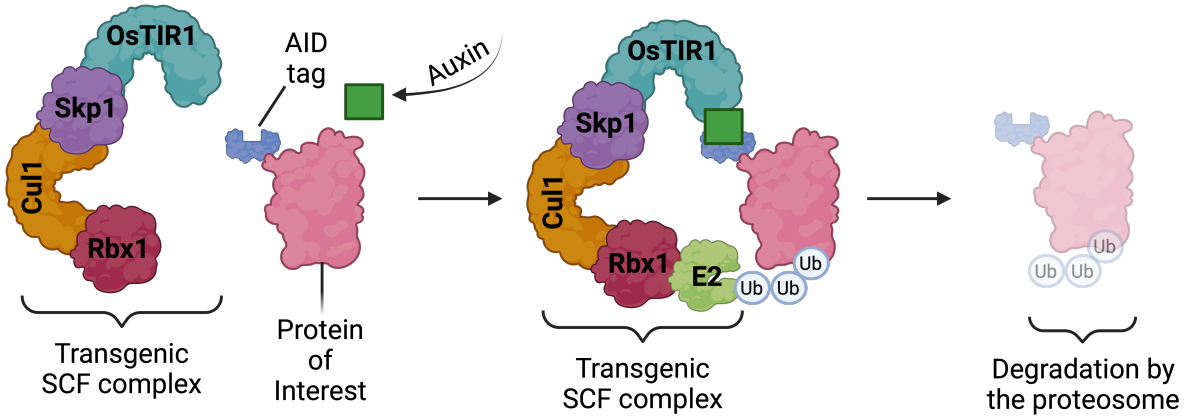



Figure 6: The auxin inducible degron (AID) system. The system requires the ectopic expression of both the POI of interest containing an AID-tag and the OsTIR1 F-box (blue). OsTIR1 forms as part of a functional transgenic SCF complex by associating with the endogenous components of the complex (Skp1, Cullin 1, and Rbx1). When auxin (green square) is introduced to the system, it binds OsTIR1 triggering recognition of the protein of interest and subsequent polyubiquitination by an E2-ubiquitin conjugating enzyme. The protein of interest is then speedily led to the proteosome for degradation. Created in BioRender.

The use of natural auxin IAA is restricted in some model organisms, and secondary effects are mitigated by using synthetic auxin, NAA (1-naphthaleneacetic acid) (Camlin and Evans, 2019). The AID system is extensively used to study the function of both essential and non-essential genes in *S. cerevisiae*, *D. melanogaster*, and *C. elegans* (Nishimura *et al.,* 2009). However, its application in human cells has been limited due to the complexity of AID-tagging endogenous proteins, especially essential ones. To address this issue, the CRISPR/Cas9 genome editing tool has been utilized to fuse the AID tag to essential genes in human cell lines (Natsume *et al.,* 2016). AID-based degradation methods have been employed to differentiate between direct and indirect transcriptional targets of transcription factors (Muhar *et al.,* 2018). Moreover, the introduction of an m-AID tag (minimum AID tag, 7-kDa) coupled with the CRISPR/Cas9 approach has been used successfully in studies involving essential human genes involved in cell division, such as APC4 and CDC7, a serine-threonine kinase involved in multiple processes including the stimulation of Aurora B activity, a key kinase required during mitosis (Lok *et al.,* 2020). In some cases, to achieve faster and more efficient POI degradation, the AID system has been combined with the Tet-OFF promoter system (Ng *et al.,* 2019). Simultaneous IAA and doxycycline treatment results in more

rapid and complete POI degradation compared to depletion mediated solely by doxycycline or IAA (Ng *et al.,* 2019). This combined system has been used to investigate essential human proteins such as CDK2, essential for S-phase progression; cyclin A, a partner of CDK1 and CDK2, involved in the control of S-phase and mitosis; and TRIP13, a regulator of the spindle assembly checkpoint (SAC) (Ng *et al.,* 2019). A limitation of the original AID system is the premature degradation of the target protein in the absence of auxin in the culture medium (Morawska and Ulrich, 2013; Nishimura and Fukagawa, 2017). This auxin-independent degradation is due to the high expression rate of TIR1. To mitigate this issue, the OsTIR1 gene can be placed under a tetracycline-regulated promoter using a Tet-Promoter in combination with the AID system (Natsume *et al.,* 2016). However, tet-OsTIR1 expression can be slow and deleteriously influence degradation timing.

In addition to the AID system, the dTAG (degradation TAG) system allows for rapid, reversible, and selective ablation of a POI (Nabet *et al.,* 2018). The dTAG strategy utilizes a modified version of the FKBP12-tag ($FKBP12^{F36V}$) and requires three components: a POI fused to $FKBP12^{F36V}$, a small heterobifunctional dTAG molecule, and the endogenous E3 ligase complex (Nabet *et al.,* 2018; Prozzillo *et al.,* 2020). The POI is fused to the 12 kDa cytosolic prolyl isomerase engineered variant $FKBP12^{F36V}$, introduced by CRISPR locus-specific knock-in. The heterobifunctional degrader (such as dTAG-13, dTAG-47 or $dTAG^{V}$-1) recruits the $FKBP12^{F36V}$-fused POI to CRBN (cereblon) – the recognition element of the CRBN-CRL4 E3 ubiquitin ligase complex – resulting in restricted degradation of the POI by the proteosome (Nabet *et al.,* 2018). The degraders each contain a moiety that acts as the synthetic selective ligand for FKBP12 (named AP1867) alongside thalidomide which binds CRBN separated by differing linker regions. More recently, the $dTAG^{V}$-1 degrader has been synthesized which engages the von Hippel-Landau (VHL) E3 ligase complex (Nabet *et al.,* 2020). This higher generation dTAG molecule eliminates the limitations of the dTAG-13 molecule in degrading several difficult-to-degrade proteins. In addition, $dTAG^{V}$-1 shows an increased pharmacodynamic and pharmacokinetic profile, exhibiting a longer half-life, improved duration of ablation, and greater exposure (Nabet *et al.,* 2020).

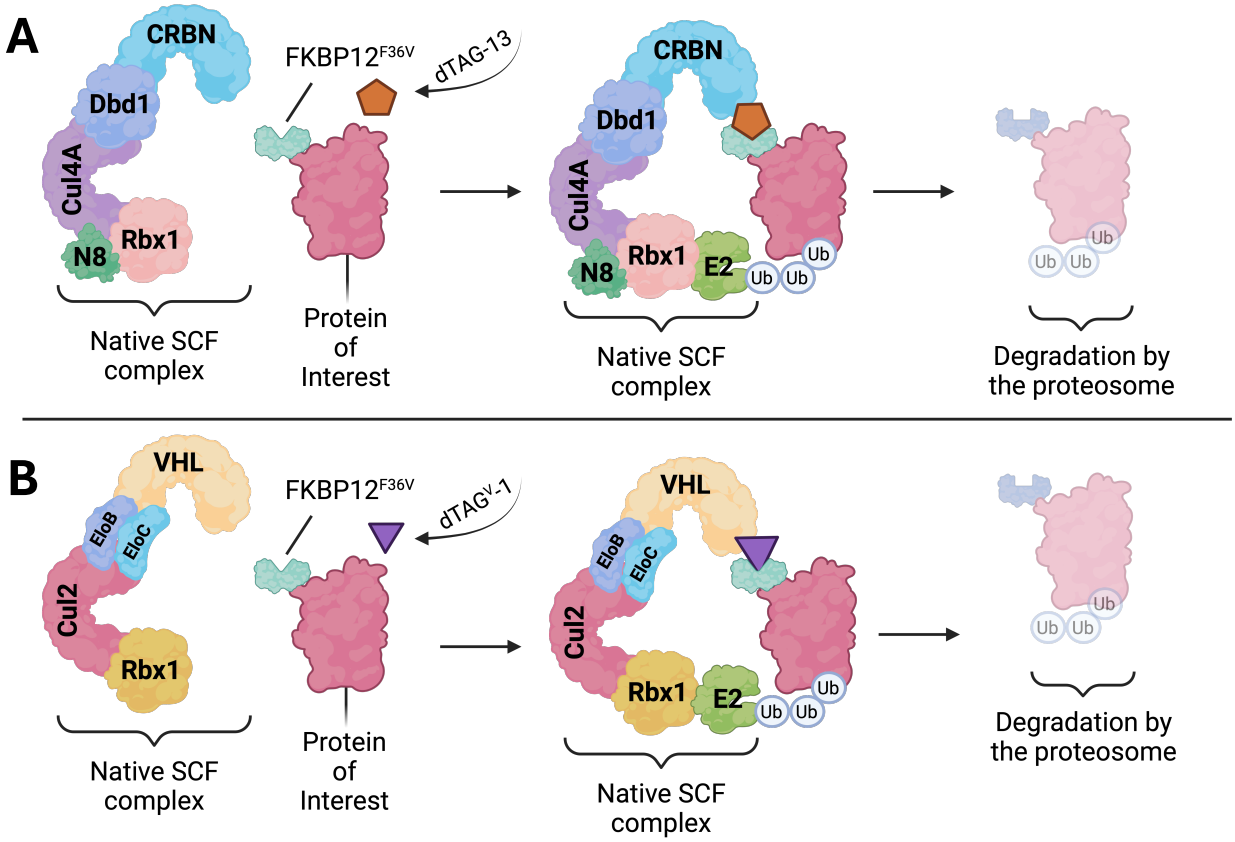



Figure 7: The degradation TAG (dTAG) system. (A) The heterobifunctional dTAG-13 molecule binds and brings together the FKBP$^{F36V}$-fused protein of interest and cereblon (CRBN), redirecting the complex towards the endogenous proteosomal machinery for degradation of the protein of interest. dTAG-13 contains AP1867 and thalidomide – ligands for FKBP12$^{F36V}$ and CRBN – respectively. CRL4-CRBN E3 ubiquitin ligase that is recruited by dTAG-13 is comprised of the substrate receptor (CRBN), adaptor protein (DBD1), cullin scaffold (CUL4A), the RING protein (RBX1) recruiting an E2 ligase and N8 ubiquitin-like protein (NEDD8). (B) The dTAG$^{V}$-1 molecule recruits the von Hippel-Landau (VHL) E3 ligase complex, which acts to increase the efficiency of the system. The VHL complex comprises the substrate receptor (VHL), two adaptor proteins (ELOB and ELOC), a cullin scaffold (CUL2) and the RING protein (RBX1). Created in BioRender.

The dTAG system was first used to evaluate the effects of acute degradation of ENL, a transcriptional regulator, and MELK, a proliferation promoting kinase, and was later employed in cells to selectively and rapidly degrade a cohort of FKBP12$^{F36V}$-fused chimeras such as KRASG12V, BRD4, HDAC1, PLK1, MYC and EZH2 (Erb *et al.*, 2017; Huang *et al.*, 2017). It was observed that dependent on subcellular compartmentalization, there exists a variation in the rate of POI degradation (Erb *et al.*, 2017; Huang *et al.*, 2017). More recently, the dTAG strategy

was employed *in vivo* using mouse models (Nabet *et al.,* 2018). Furthermore, the technology has been leveraged to rapidly ablate all protein IE2 isoforms to identify the role they play in late human cytomegalovirus infection, and additionally to degrade solute carrier proteins (SLC proteins) – the largest class of transporters with multi-pass transmembrane domain topology (Bensimon *et al.*, 2020; Li *et al.,* 2020). Overall, the dTAG system is a flexible, universally applicable, selective ablation strategy absent of significant off-target effects.

## Research Aims

INTS3 is a constituent of the Integrator complex, which functions to mediate promoter-proximal pausing in addition a variety of other roles, and has been implicated in dsDNA damage repair in conjugation with the SOSS complex. However, the mechanism by which INTS3 contributes to overall transcriptional regulation remains elusive. It is unknown if INTS3 plays a direct role, or rather occupies a more accessory position, in the process of transcriptional regulation. This work aims to elucidate the multifaceted roles of INTS3 in transcriptional regulation. Firstly, this work aims investigate if INTS3 can be rapidly and reliably ablated by targeted protein degradation and hence probe the short-time scale downstream effects of such degradation on the composition, expression and function of the Integrator complex and RNAPII. Furthermore, this work aims to explore the effects INTS3 has on the transcriptome in terms of control of the up- or down-regulation of particular subsets of genes. Moreover, the potential presence of the SOSS complex alongside Integrator across the genome during transcription will be investigated. Lastly, the effects of INTS3 ablation on RNAPII gene occupancy and processivity will be explored.

# *Chapter 2: Results*

**An INTS3 dTAG system was generated for rapid and reliable ablation of INTS3 at the protein level**

Targeted protein degradation (TPD) strategies are attractive approaches to degrade a protein of interest (POI) by virtue of their rapidity of protein ablation, selectivity, reversibility and absence of off-target effects. In this light, a dTAG TPD was engineered in OVCAR8 cells to probe the effects of INTS3 on transcriptional regulation in a broader sense. The initial step to designing such a system involves the generation of two plasmids. The first is an sgRNA plasmid which encodes a CRISPR guide RNA (gRNA) of 20 bp (excluding the PAM sequence 5`-NGG-3`) and V2.0 spCas9 enzyme to perform precise DNA cutting within 10 – 30 bp of the insert. The second is a homology-directed repair (HDR) plasmid which contains a left and right homology arm of 1 kb length flanking the desired insert: in this case, a sequence encoding a linker $(GGGGS)_3$ region, the cytosolic prolyl isomerase variant $FKBP12^{F36V}$, two HA-tags, a P2A ribosomal skipping sequence, and an antibiotic resistance cassette (Figure 8). The left-homology arm (LHA) of the HDR plasmids was designed to incorporate synonymous mutations to prevent re-cutting of spCas9 after initial cleavage through alteration of the pre-PAM recognition sequence once recombination had occurred (Figure 8). Such synonymous mutations were decided upon based on codon adaption index (CAI) similarity. The right homology arm (RHA) was amplified from extracted gDNA from OVCAR8 cells to maintain sequence isogeneity. Two sgRNA plasmids targeting the C-terminus of INTS3 (exon 30) (Figure 9) and three HDR plasmids each with differing antibiotic resistance cassette (puromycin, blasticidin and neomycin) (Figure 10) were generated by molecular cloning and / or Gibson assembly. Each sgRNA plasmid was sequenced by Sanger sequencing to ensure accurate assembly and seamless joining between gRNA and the plasmid backbone (Figures 9C and 9D).

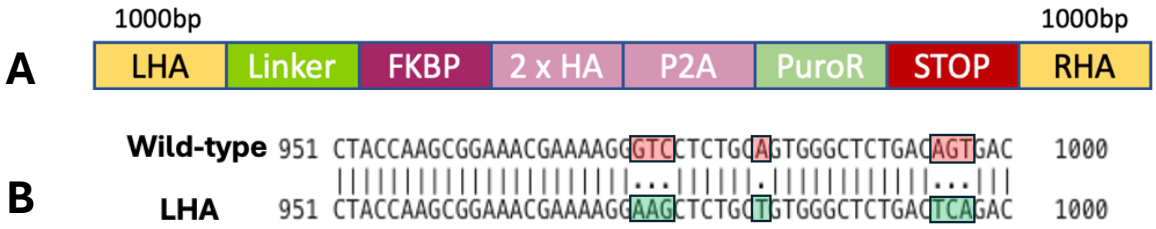



Figure 8: Schematic of insert for HDR plasmids and synonymous mutations present in LHA. (A) The HDR plasmids were generated to have a left homology arm (LHA) of 1 kb length with synonymous mutations followed by a $(GGGGS)_3$ linker, the cytosolic prolyl isomerase variant $FKBP12^{F36V}$, two HA-tags, a P2A ribosomal skipping sequence, and an antibiotic resistance cassette (shown: puromycin resistance). The right homology arm (RHA) is likewise 1 kb in length consisting of the stop codon and downstream 3` sequence from the INTS3 gene. (B) Synonymous mutations were included in the LHA of the HDR plasmids to prevent re-cutting post-cleavage by spCas9; red indicates wild-type sequence while green indicates the synonymous mutation.

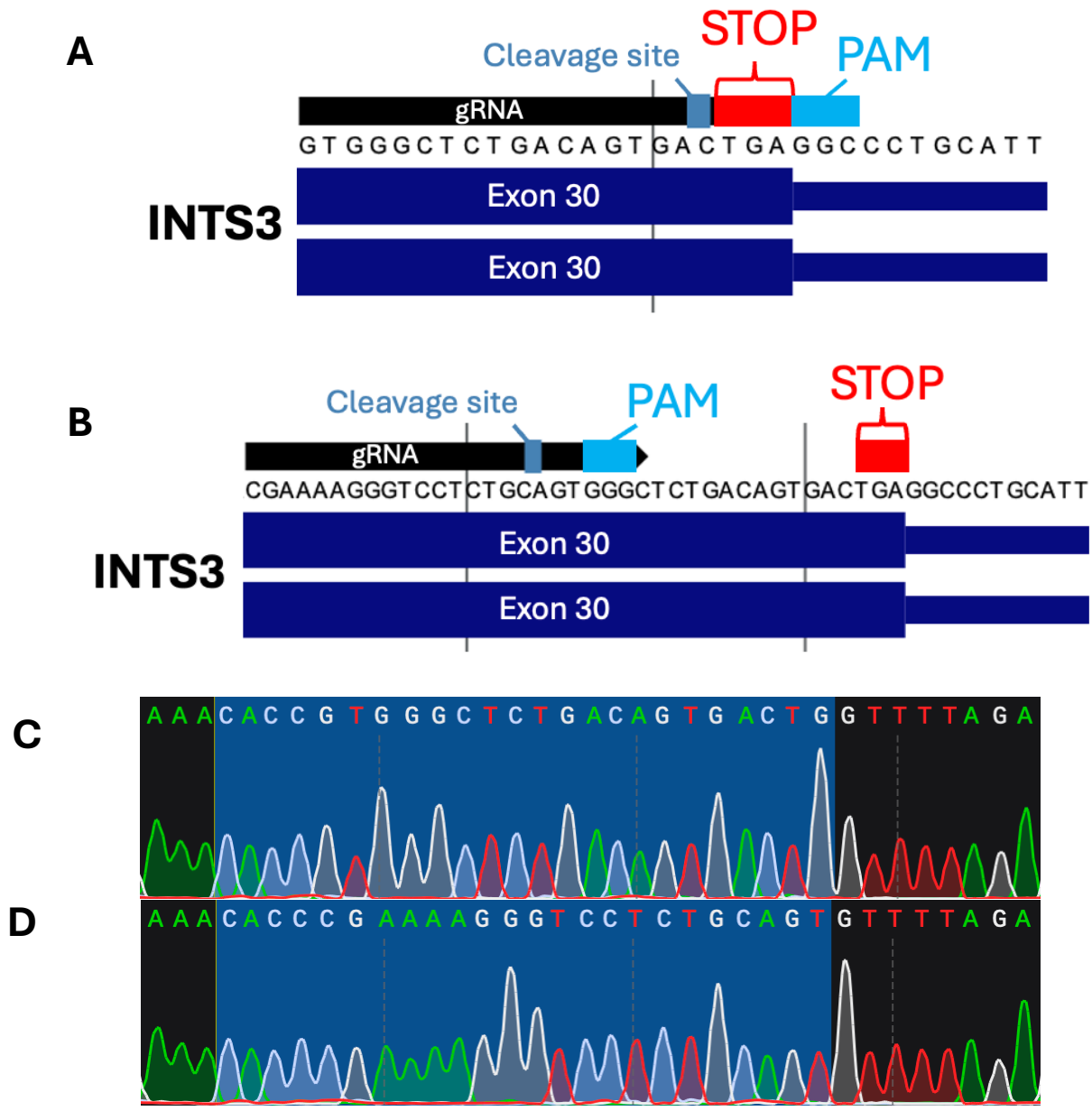



Figure 9: Schematic of gRNA sequences to cleavage sites along INTS3. Two sgRNA plasmids were designed to each incorporate one of two gRNA for cleavage within 30 bp of the stop codon. (A) The first sequence gRNA allows for cleavage between A/C on the sense strand, within 4 bp of the stop codon and comprises the recognition sequence 5`-**GTG GGC TCT GAC AGT GAC TG**-3`. (B) The second sequence gRNA allows for cleavage between C/A of the sense strand, within 22 bp of the stop codon and comprises the recognition sequence 5`-**CGA AAA GGG TCC TCT GCA GT**-3`. (C and D) Sanger sequencing confirmed the gRNA sequence was seamlessly inserted into the sgRNA delivery plasmid.

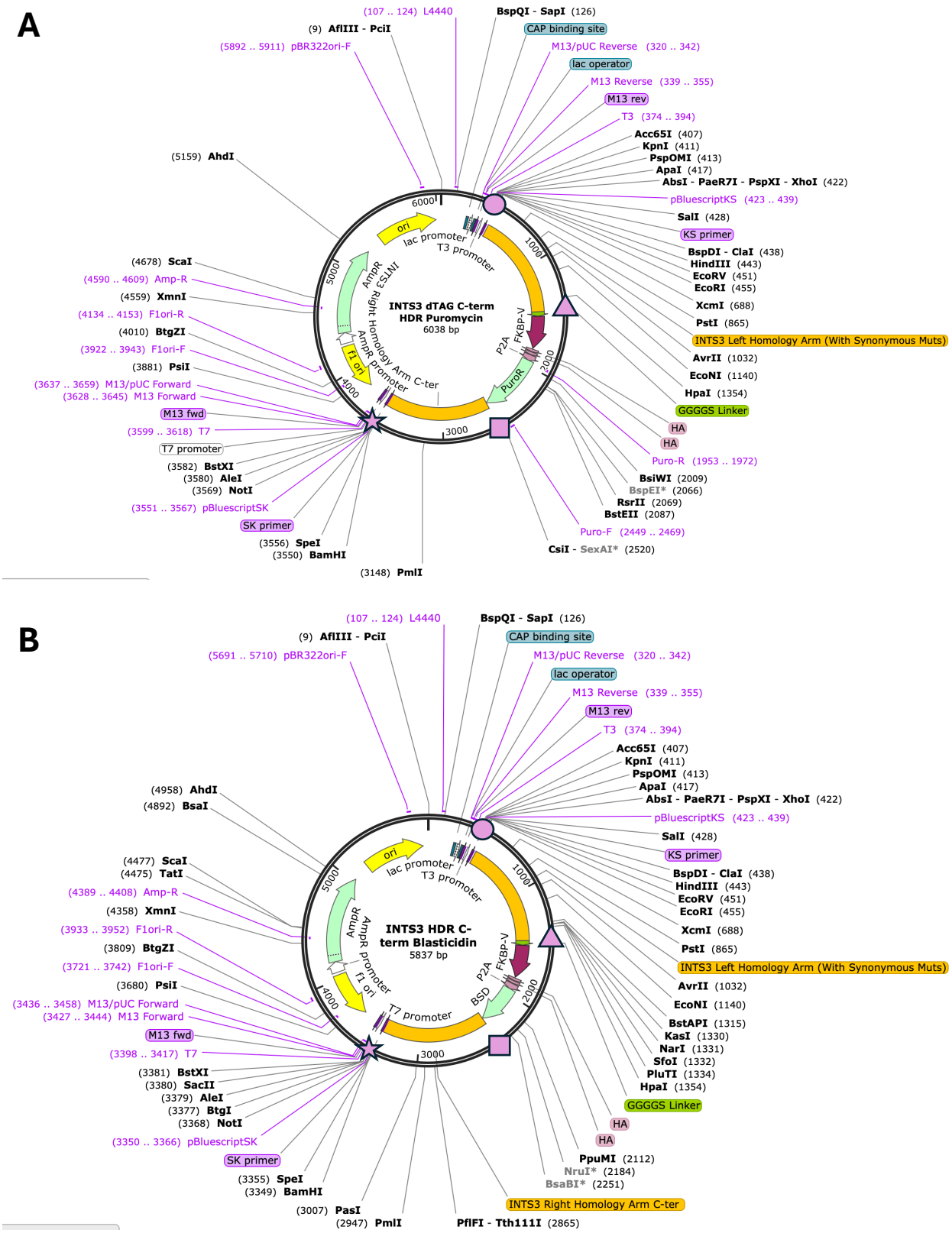



Figure 10 (continued on following page)

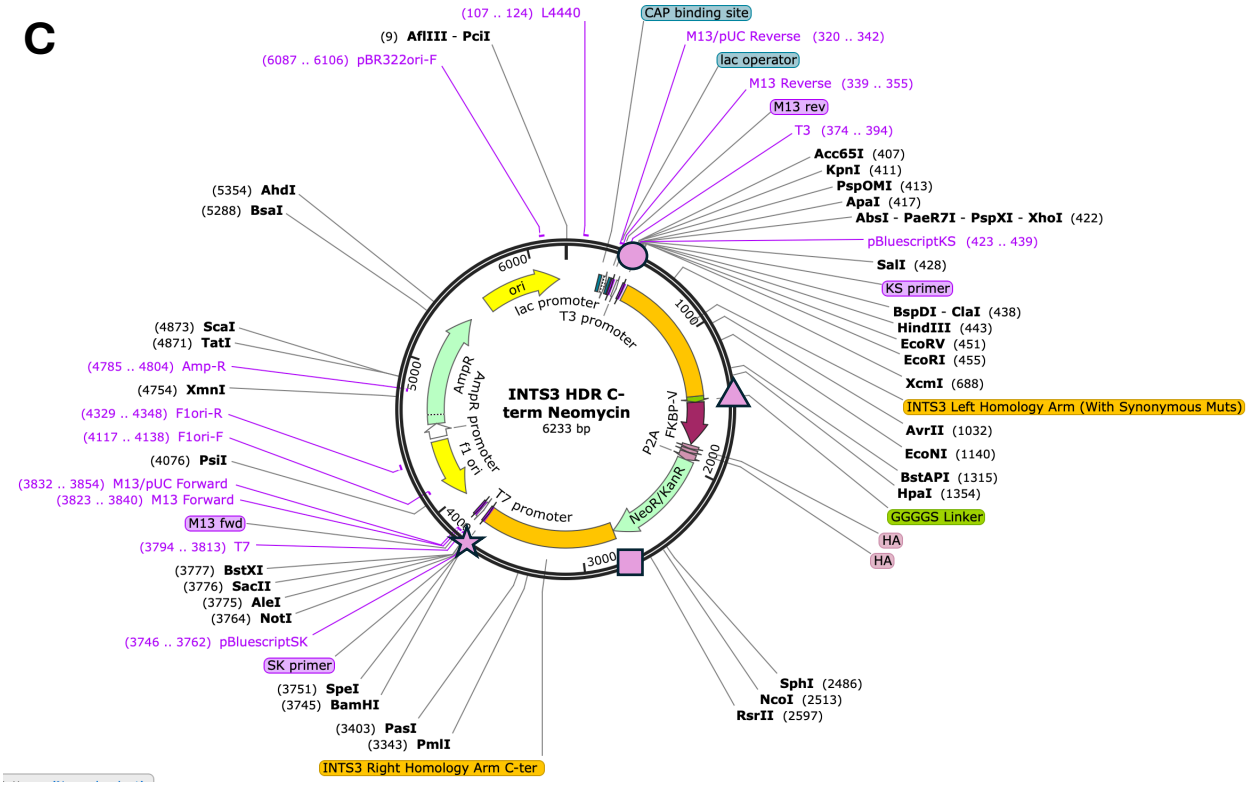



Figure 10: HDR plasmids constructed for insertion of FKBP12$^{F36V}$ by homology directed repair. (A) The HDR plasmid was designed to have a pBlueScript backbone (segment between star and circle); a LHA synthetically synthesized to be homologous to 1000 bp upstream of the codon stop site (orange, labelled "INTS3 Left Homology Arm (with Synonymous Muts)") (segment between circle and triangle); an insert containing a linker $(GGGGS)_3$ region, the cytosolic prolyl isomerase variant FKBP12$^{F36V}$, two HA-tags, a P2A ribosomal skipping sequence, and an antibiotic resistance cassette (segment between triangle and square); and the RHA which was amplified from OVCAR8 gDNA to maintain isogeneity (orange, labelled "INTS3 Right Homology Arm C-ter") (segment between square and star). HDR plasmids either incorporated an antibiotic resistance cassette for puromycin (labelled PuroR, green), (B) blasticidin (labelled BSD, green) or (C) neomycin (labelled NeoR/KanR, green).

Two positive clones (C1 and C4) for the initial knock-in were generated through electroporation of the either of the sgRNA alongside the HDR plasmid containing puromycin resistance into OVCAR8 cells (Figure 11A). The CRISPR/Cas9 knock-in was validated by PCR amplification of gDNA for the region surrounding the knock-in site. The expected size of the amplicon was 1293 bp. Two high-fidelity polymerases (Q5 and Phusion) were employed to confirm the knock-in was successful; it was observed that both clones were homozygous for the insert knock-in in comparison to the negative control (Figure 11B and 11C).

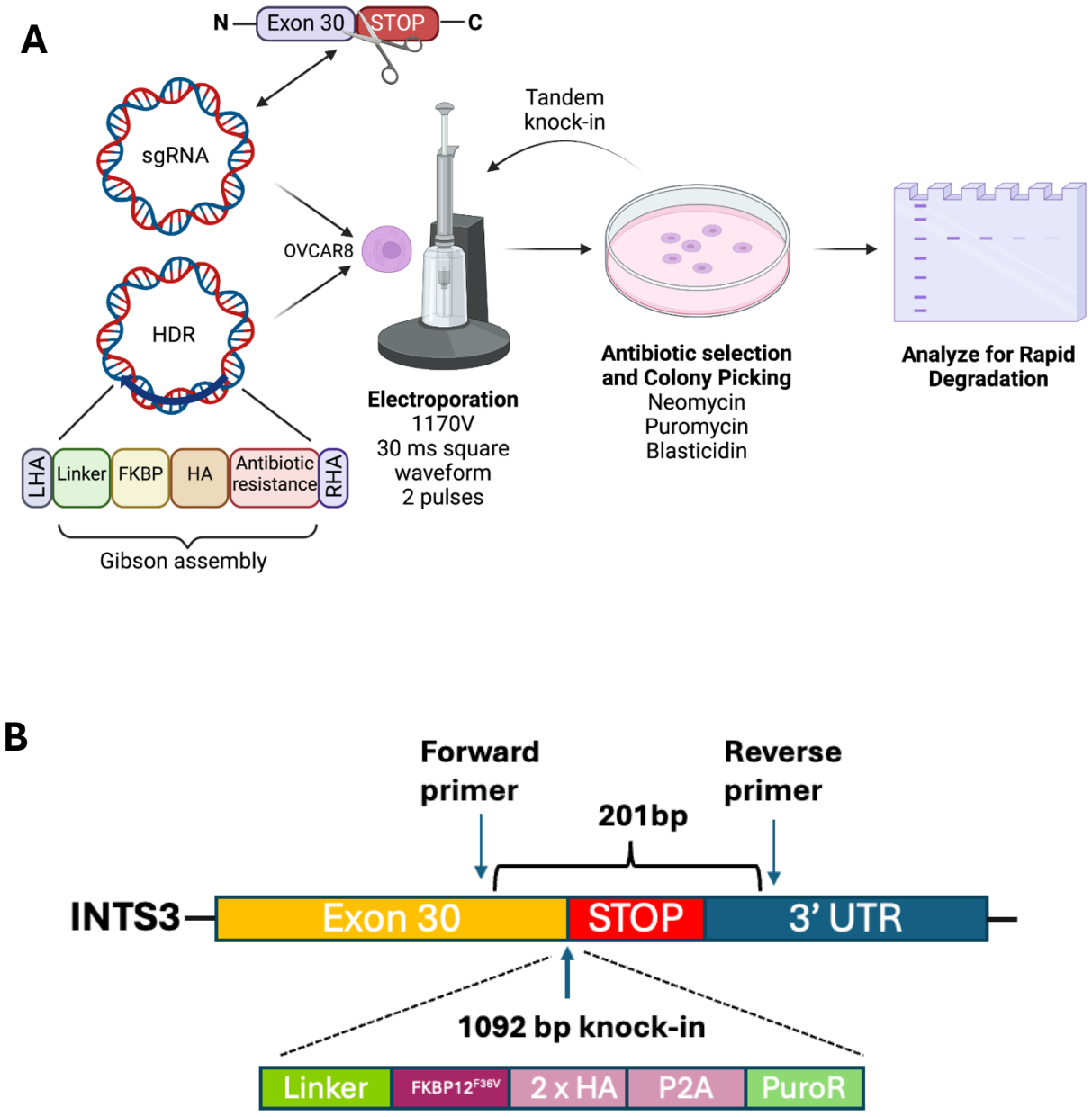



Figure 11 (continued on following page)

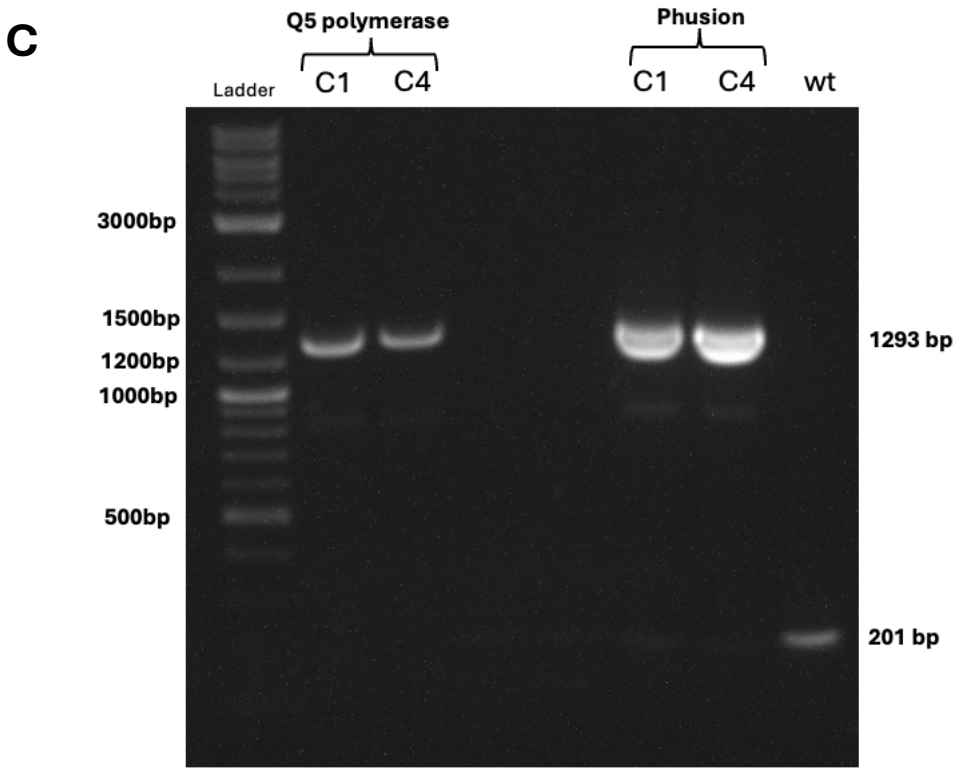



Figure 11: CRISPR/Cas9 knock-in validation by PCR if gDNA from OVCAR8 cells. (A) Schematic of the electroporation procedure for transfection of OVCAR8 cells with both the sgRNA and HDR plasmid containing puromycin resistance. OVCAR8 cells were electroporated at 1 170 V, with a 30 ms square waveform for two pulses, followed by selection under antibiotic and eventual probing for INTS3 degradation by immunoblotting (B) Schematic of the knock-in amplification strategy to validate insertion. The insert containing the linker $(GGGGS)_3$ region, the cytosolic prolyl isomerase variant $FKBP12^{F36V}$, two HA-tags, a P2A ribosomal skipping sequence, and an antibiotic resistance cassette had a length of 1092 bp while the region around the knock-in for amplification was 201 bp in length. Amplification was conducted around the end of exon 30 at the C-terminal end of the *INTS3* gene and the 3`-untranslated region of the *INTS3* gene. (C) Agarose gel electrophoresis resolved the insert was present within the amplified region confirmed by two polymerases: Q5 and Phusion. Clones C1 and C4 were homozygous positive for an insert of size 1293 bp confirmed by both polymerases. Amplification of wt gDNA showed an amplicon of 201 bp serving as a negative control.

The first attempt at INTS3 ablation by dTAG treatment was confirmed by Western blot. Degradation was conducted over a long-series time course, from no treatment of dTAG molecule ($dTAG^V$-1) up to eight hours of treatment in two-hour increments. Degradation was monitored by immunoblotting for both INTS3 and HA-tag. While it was observed that the HA-tagged INTS3 is present in both clones C1 and C4, and subsequently degrades within two hours of dTAG treatment with complete degradation maintained up to the eight hour time point, INTS3 levels did not seem

to change; this was evident by a signal for each of the time-course increments equivalent to that of the untreated samples for both clones (Figure 12).

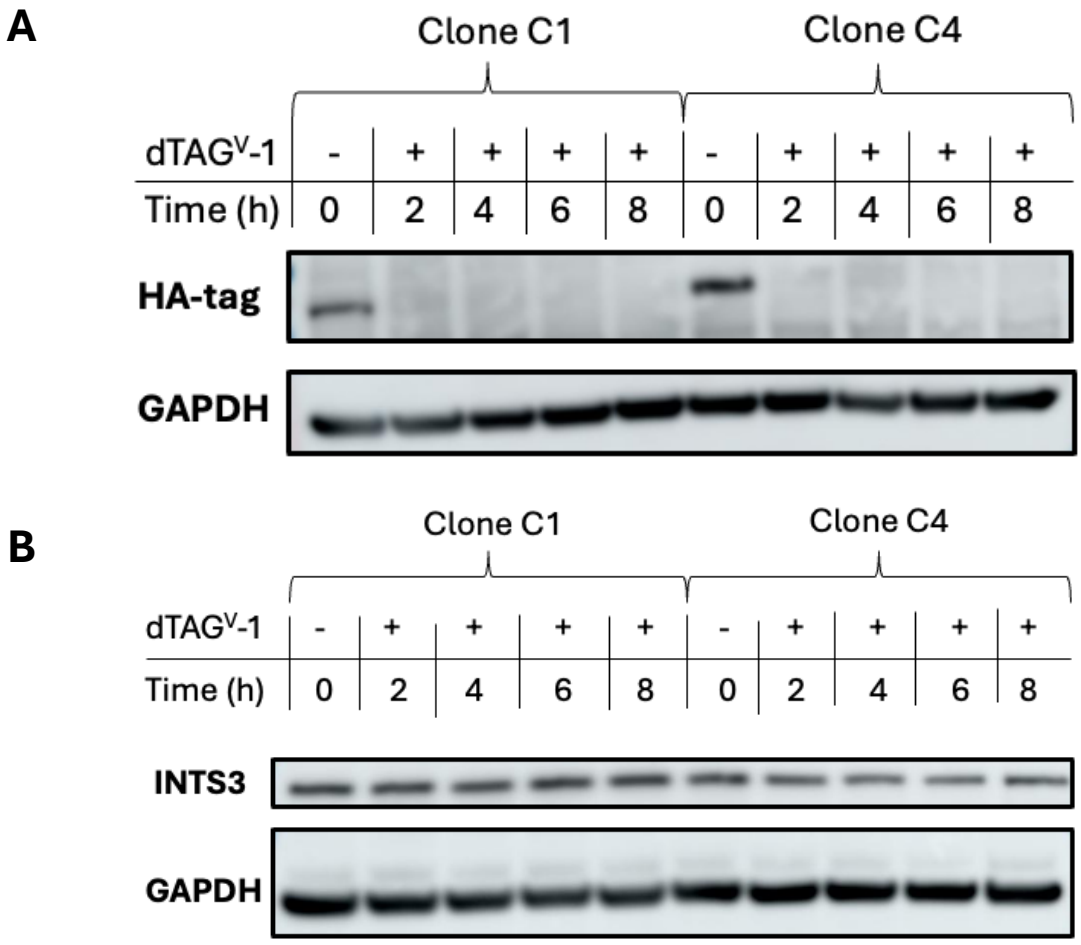



Figure 12: Immunoblots for dTAG induced degradation for clones C1 and C4. (A) The HA-tag inserted via CRISPR/Cas9 knock-in showed a signal for complete degradation within two hours of dTAG treatment and was maintained for up to eight hours of treatment with dTAG. (B) INTS3 did not show a signal indicative of degradation for any time point up to eight hours.

It was suspected that the reason for degradation to be observed for the HA-tag but not for INTS3 was hyperploidy of chromosome 1 in OVCAR8 cells, thus leading to an incomplete knock-in of the insert into a subset of *INTS3* – but not into each copy of the *INTS3* gene. In this light, electroporation was repeated however with two major alterations: firstly, electroporation was performed on clone C4 as opposed to wtOVCAR8 to increment on any knock-in already present in the system and, secondly, using an HDR plasmid with a different antibiotic resistance cassette

(but the same insert containing the aforementioned elements). Electroporation was once again performed according to the same parameters as originally used using either of the sgRNA plasmids and the HDR plasmid containing puromycin resistance. Cells were then selected for under antibiotic selection pressure of both puromycin and blasticidin, and positive colonies identified for screening by Western blot. Only one clone (C20) was truly positive for INTS3 ablation. Immunoblotting was performed for both a long-time course (0 - 12 hours with two- or six-hour increments) and a short-time course (0 – 120 minutes, with 15-to-60-minute increments). For the long-time course, an INTS3 degradation signal was observed whereby for 2 hours and up to 12 hours, the signal for INTS3 was seen to decrease by approximately 80 – 90 % (Figure 13). However, this was not observed for the short-time course: INTS3 signal remained persistent from 0 – 120 minutes despite the signal for the HA-tag showing signature complete degradation within 15 minutes of dTAG treatment (Figure 13).

Electroporation was once again repeated using the same parameters on clone C20 to increment on any knock-in already present in the system using a third HDR plasmid with a different antibiotic resistance cassette (neomycin). Following electroporation, cells were placed under selection pressure of all three antibiotics conferring resistance, namely, puromycin, blasticidin and neomycin (G418 sulphate). As it is more crucial for short-term degradation to be observed in the system (within two hours of dTAG treatment), a short-time course was performed at zero hours and two hours of treatment, followed by an extended short-time course at 0 – 120 minutes with 15-to-60-minute increments. Immunoblotting for five clones exhibited a signal indicative of INTS3 degradation (Figure 14). In addition, immunoblotting for one of the five positive clones (clone 10) showed a signal indicative of short-time course degradation from 0 – 120 minutes, with degradation observed to be approximately 80 – 90 % after one to two hours dTAG treatment, with approximately 60 % degradation within 30 minutes of treatment (Figure 14). Two other clones (clones 3 and 16) were screened by short-time course immunoblot but did not show the same robustness of degradation as observed for clone 10 (Figure 14), showing degradation of approximately 50 % and 70 %, respectively, after two hours of dTAG treatment in addition to approximately 40 % degradation within 30 minutes. Clone 10 was, hence, amplified, stored, and used for all subsequent downstream experiments.

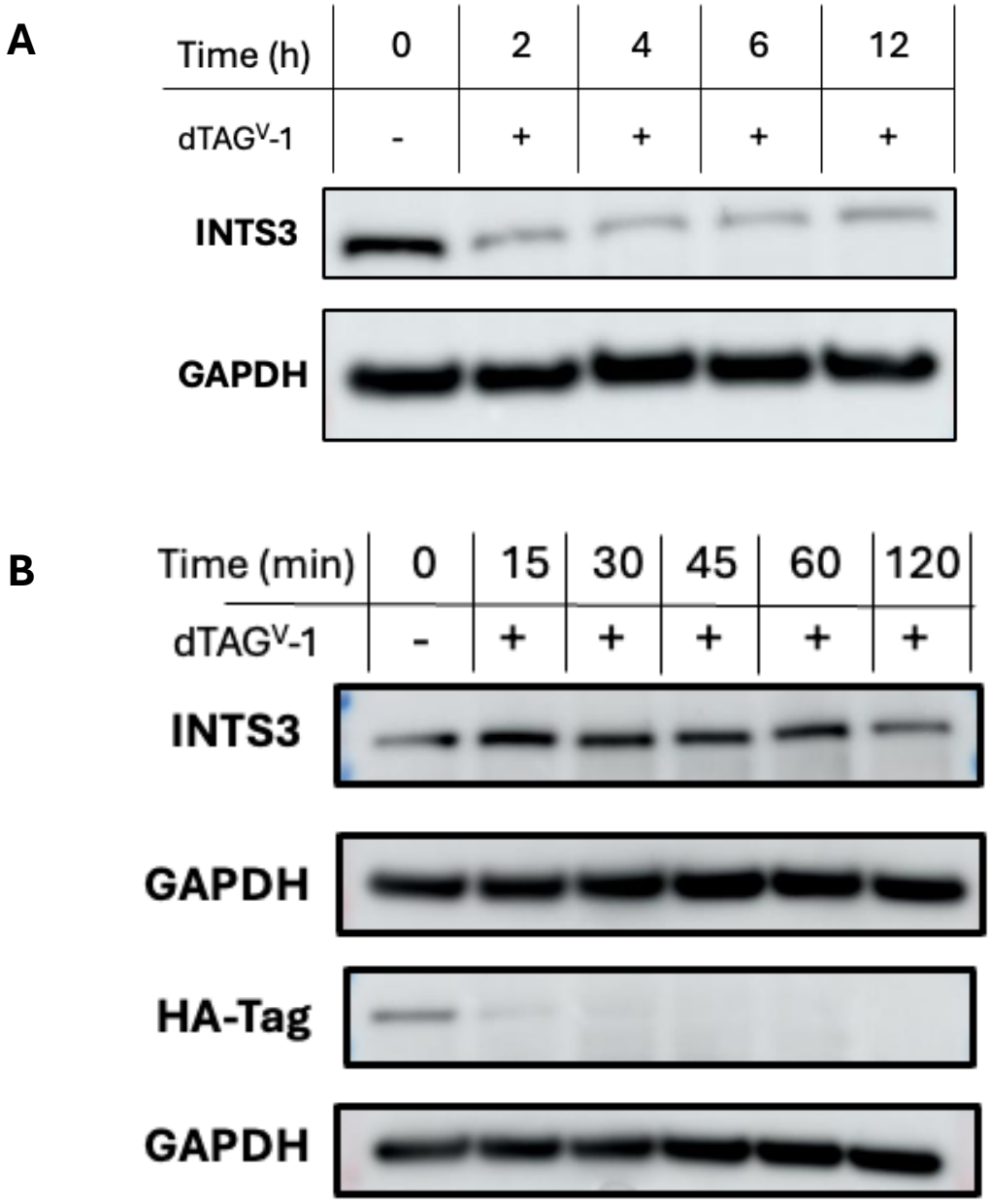



Figure 13: Western blot for long- and short-time course of INTS3 degradation by dTAG treatment. (A) Over 0 – 12 hours of treatment, a signal was observed for INTS3 degradation of approximately 80 – 90 % decrease in signal. Degradation was maintained throughout the long-time course from 2 – 12 hours. (B) Despite a signal indicative of degradation of INTS3 was observed for the long-time course, over a short-time course INTS3 was not observed to degrade from 0 – 120 minutes of dTAG treatment despite the HA-tag showing a positive degradation signal over the same time period.

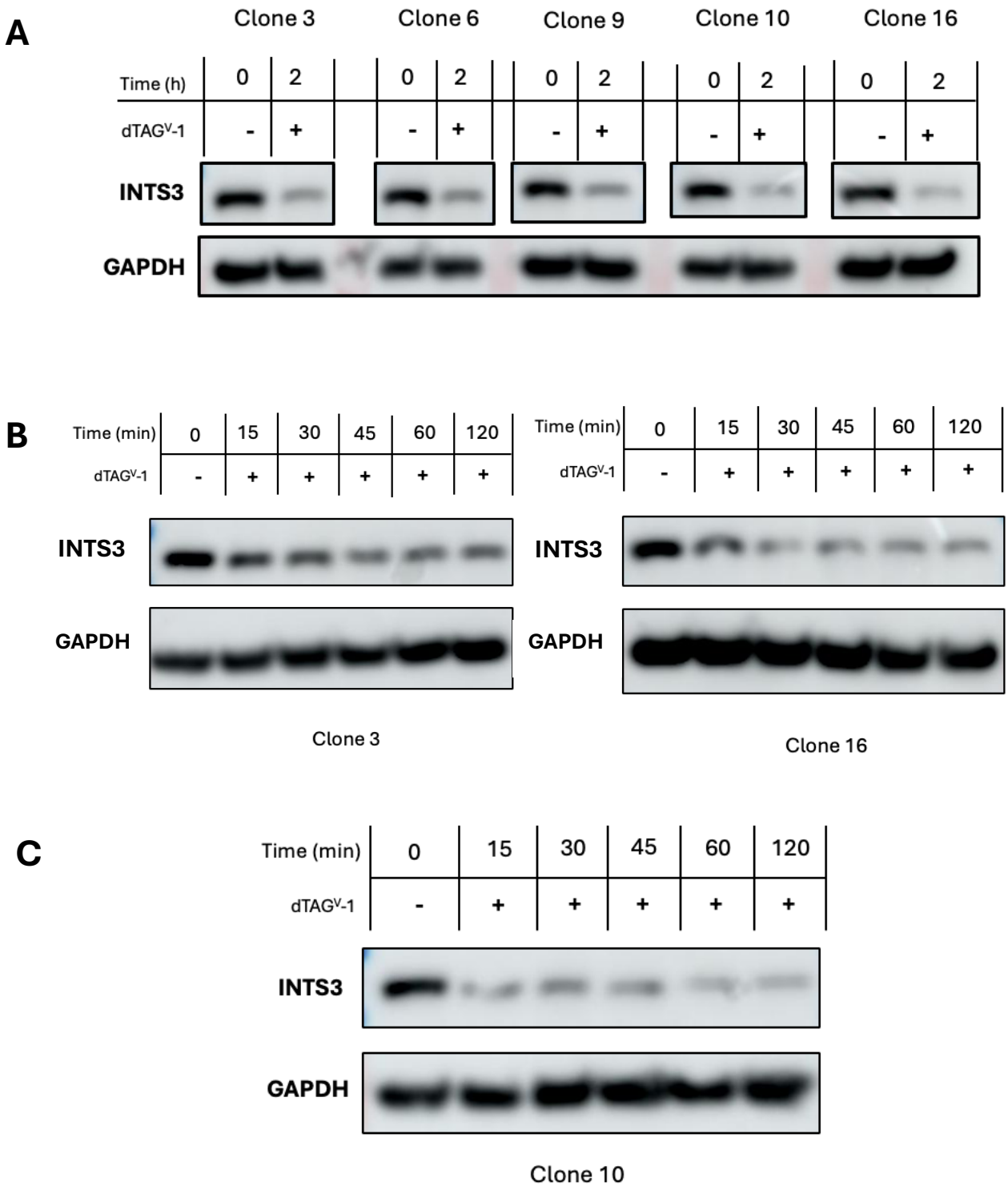



Figure 14: Short-time course of INTS3 degradation after treatment with dTAG for clones 3, 16, and 10. (A) Five clones showed a positive signal for degradation after dTAG treatment for two hours. (B) Clones 3 and 16 showed a positive signal for degradation with varying degrees of robustness; clone 3 appeared to show degradation in two hours of approximately 50 % while clone 16 showed degradation in two hours of approximately 70 %. (C) Clone 10 showed the

most robust degradation when monitored by short-time course with approximately 90 % degradation of signal observed after one to two hours, with approximately 60 % of signal degraded after 30 minutes of dTAG treatment.

INTS3 is a component of the SOSS complex, comprised of NABP1/2 and INIP, which senses ssDNA and participates in DNA damage repair. In order to validate that INTS3 was being ablated by dTAG treatment, co-immunoprecipitation of INTS3 and SOSS complex components was performed. In addition, to observe if INTS3 associated with the phosphatase complex of Integrator, co-immunoprecipitation for PPP2R1A (a PP2A-A subunit) was performed. Clone 10 cells were treated for an hour with $dTAG^{V}$-1 followed by co-immunoprecipitation for the HA-tag – incorporated into the C-terminus of INTS3 – and immunoblotted for respective antibodies. INTS3 was observed to be present in the input and was ablated within one hour of dTAG treatment (Figure 15). INTS3 was pulled down in the untreated sample but not for the dTAG treated sample, further indicating its ablation (Figure 15). PPP2R1A, a constituent of Integrator, was present in the input and treated sample but did not co-immunoprecipitate with INTS3 (Figure 15). NABP2, a member of the SOSS complex, was present in both the input and dTAG treated samples but only co-immunoprecipitated in the untreated sample as opposed to the dTAG treated sample (Figure 15). As INTS3 acts as a scaffold for the SOSS complex, allowing for NABP1/2 to dock to its N-terminal cleft, the absence of INTS3 should prohibit co-immunoprecipitation of NABP1/2. The absence of NABP2 co-immunoprecipitating with the dTAG treated sample is indicative of INTS3 ablation after dTAG treatment, further confirming the robustness of the INTS3 dTAG TPD system (Figure 15).

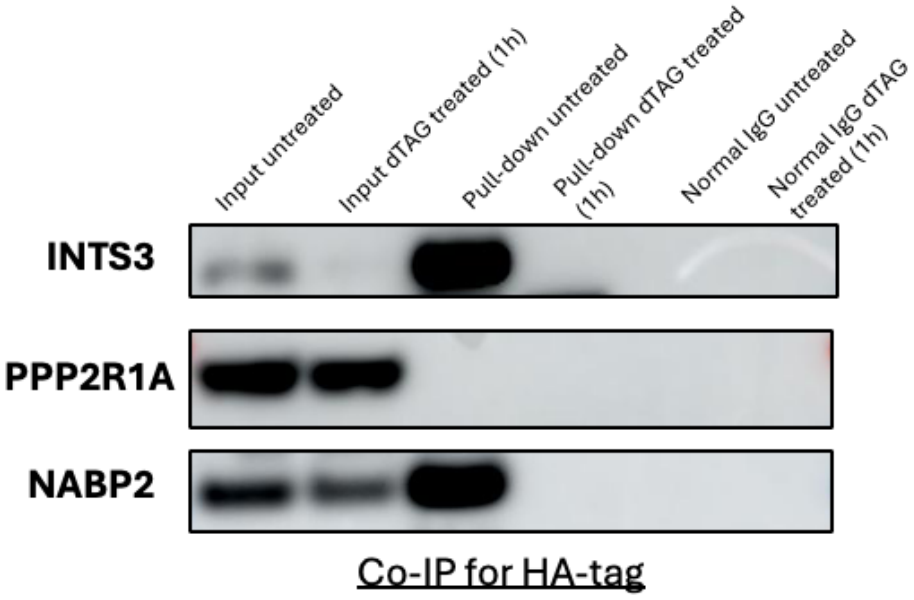



Figure 15: Co-immunoprecipitation pull down of HA-tag for clone 10. INTS3, PPP2R1A and NABP2 were observed to be present in the input. INTS3 was showed a signal indicative of ablation after dTAG treatment for one hour while signals for PPP2R1A and NABP2 were unaffected by dTAG treatment. Following co-immunoprecipitation by HA-tag and immunoblotting for the respective antibodies, INTS3 and NABP2 were pulled down in the untreated samples, although PPP2R1A was not pulled down. For the dTAG treated samples, neither INTS3, NABP2 or PPP2R1A were pulled down.

**Integrator subunits and module expression, and CTD phosphoserine state, is affected by INTS3 ablation**

Integrator is comprised of at least 15 subunits clustered into discrete functional modules which confer activity to the complex. Such functional modules include the endonucleolytic module (INTS11, INTS9 and INTS4), the enhancer/arm module (INTS10, INTS13, INTS14 and INTS15), the scaffold module (INTS1, INTS2, and INTS7), and the phosphatase module (INTS8, INTS6, INTS3, and PP2A (as a part of INTAC)). In order to assess the effect of INTS3 ablation on Integrator subunits and the complex formation as a whole, clone 10 cells were treated for four hours with dTAG$^{V}$-1 treatment and the expression of various Integrator subunits probed by Western blot. It was observed that various Integrator subunits were upregulated post-treatment with the dTAG molecule alongside the phosphoserine states of the CTD of RNA Pol II, all the while INTS3 was observed to be degraded. Specifically, INTS6 of the phosphatase module, INTS1 of the scaffold module, and INTS5 were observed to be upregulated, while INTS11 of the

endonucleolytic module and INTS10 of the enhancer module did not show a measurable change in signal (Figure 16). Furthermore, the phosphoserine states of Ser2, Ser5 and Ser7 of the CTD of RNA Pol II were observed to be more phosphorylated after dTAG treatment for four hours while levels of RNA Pol II showed a constant signal regardless of dTAG treatment (Figure 16). PPP2R1A and PPP2CA, the scaffolding and catalytic subunits of PP2A, respectively, did not show any observable change after dTAG treatment (Figure 16). Moreover, NABP1of the SOSS complex did show an observable increase in signal post-treatment (Figure 16). Retroactively, EGR1 was also probed and was observed to be upregulated post-treatment (Figure 16); the rationale behind probing for EGR1 will be made evident later on in the work.

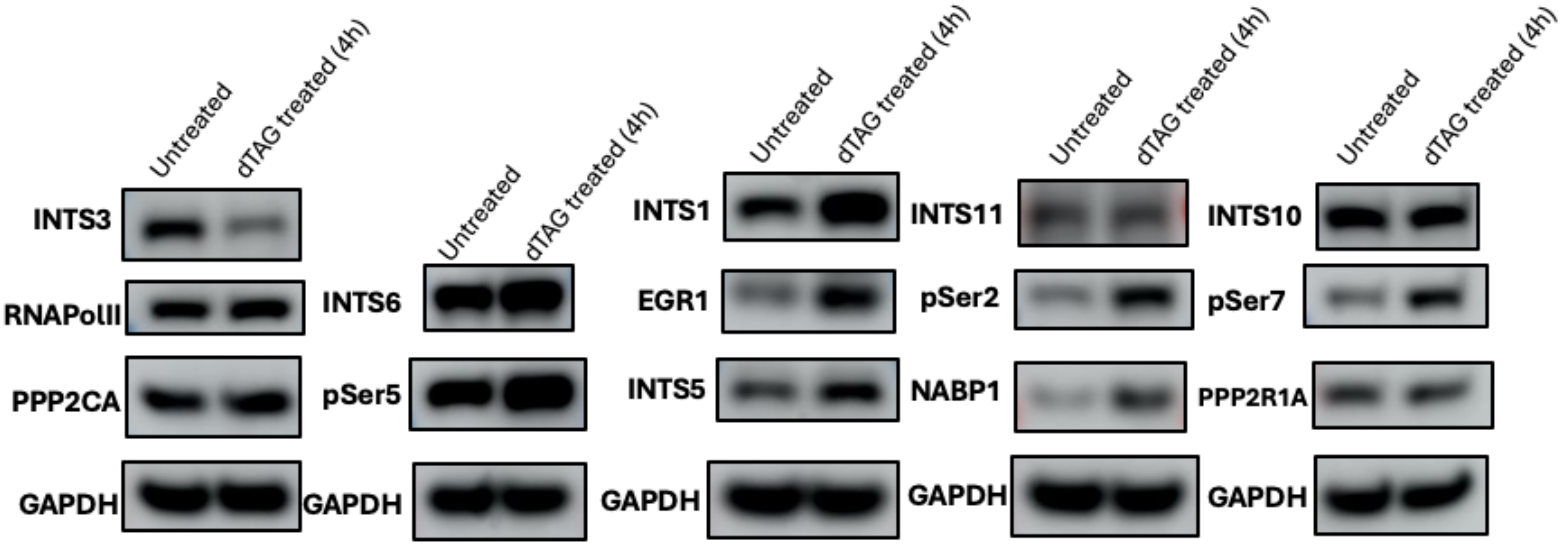



Figure 16: Integrator, phosphoserines of the CTD of RNA Pol II, PP2A, NABP1 and EGR1 change in signal detected by Western blot post-dTAG treatment for four hours. INTS3 was observed to be degraded after four hours dTAG treatment. INTS6 of the phosphatase module, INTS1 of the scaffold module, and INTS5 were observed to be upregulated, whereas INTS11 of the endonucleolytic module and INTS10 of the enhancer module showed no measurable change in signal. Additionally, the phosphoserine states of Ser2, Ser5, and Ser7 of the CTD of RNA Pol II were more phosphorylated after four hours of dTAG treatment, while RNA Pol II levels remained constant regardless of dTAG treatment. PPP2R1A and PPP2CA, the scaffolding and catalytic subunits of PP2A, respectively, showed no observable change post-dTAG treatment. Conversely, NABP1 of the SOSS complex exhibited an increase in signal following treatment. Retrospectively, EGR1 was also probed and found to be upregulated post-treatment.

The same experiment was repeated for dTAG treatment for eight hours to observe if changes to Integrator subunits, phosphoserine states, NABP1, PP2A and EGR1 were maintained. As expected, INTS3 was degraded within eight hours of dTAG treatment (Figure 17). However, the effects of upregulation of subunits INTS6, INTS1 and INTS5 were more pronounced (Figure 17). In particular, INTS6 showed significant increase of signal for the dTAG treated sample in comparison to the dTAG untreated sample. INTS11 and INTS10 were unchanged in terms of signal for dTAG untreated against treated. The degree of CTD serine phosphorylation was also more conspicuous after eight hours dTAG treatment. In particular, pSer5 showed a substantial increase in signal for the dTAG treated sample as compared to the untreated sample, with pSer2 and pSer7 showing modest increases in phosphorylation after treatment (Figure 17). Unlike after four hours dTAG treatment, RNA Pol II also showed an increased signal for dTAG treated against untreated. Unlike after treatment for four hours, PPP2CA showed a modest increase in signal after dTAG treatment for eight hours while PPP2R1A showed no change (Figure 17). Similar to treatment for four hours, NABP1 and EGR1 also showed an increase in signal for the treated sample as opposed to the untreated sample.

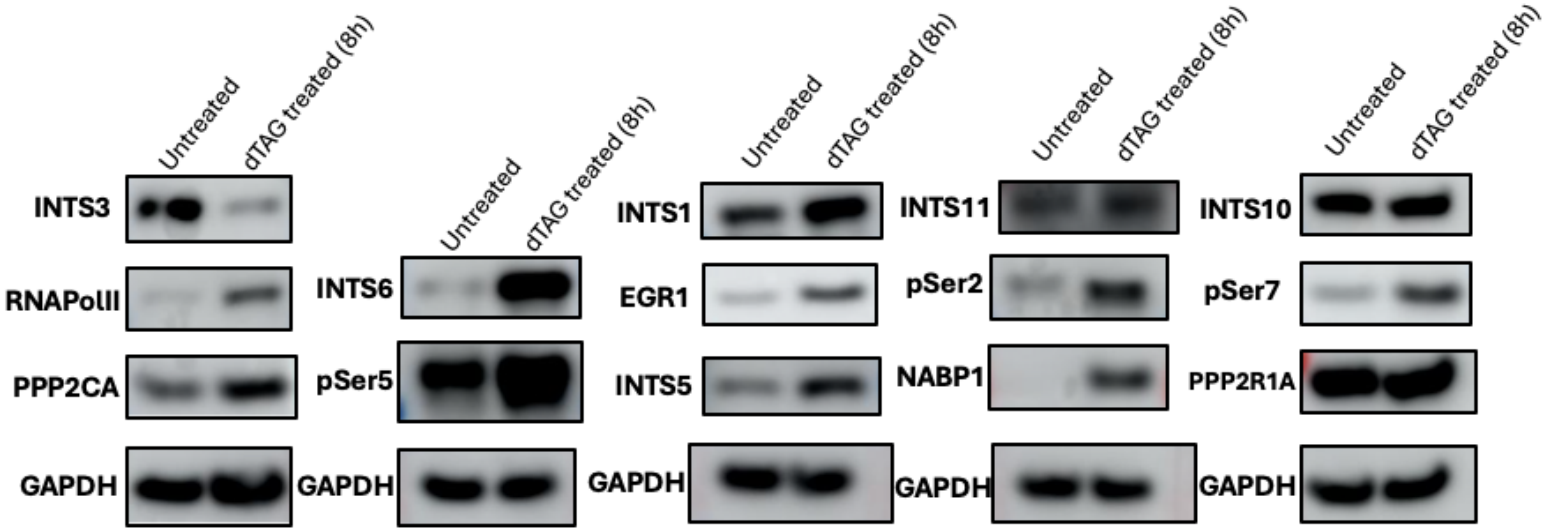



Figure 17: Integrator, phosphoserines of the CTD of RNA Pol II, PP2A, NABP1 and EGR1 change in signal detected by Western blot post-dTAG treatment for eight hours. INTS3 was observed to be degraded after eight hours dTAG treatment. Integrator subunits INTS6, INTS1 and INTS5 were observed to have an increased signal after dTAG treatment in comparison to the untreated sample, with INTS6 showing a significant increase in signal. RNA Pol II, pSer2, pSer5 and pSer7 additionally showed an increase in signal with pSer5 showing a substantial increase of signal for dTAG treated versus untreated. PPP2CA was observed to have a slight increase in signal, however, PPP2R1A showed no difference in signal for untreated compared to dTAG

treated for eight hours. NABP1 and EGR1 furthermore were observed to have an increase in signal of treated against untreated, whereby NABP1 significantly showed an increase in signal.

Given the results of upregulation of various Integrator subunits following dTAG treatment, a glycerol gradient was performed to assess if any changes to the macromolecular assembly of the Integrator complex was accompanied by INTS3 ablation. Nuclear extracts of clone 10 cells were collected for cells untreated and treated for four hours by dTAG treatment and separated by isopycnic centrifugation along a glycerol gradient of 11 – 50 % glycerol. INTS1, PPP2R1A, NABP1, NABP2, RNA Pol II and PPP2CA were probed (Figure 18). For the untreated samples, INTS1 was observed to fractionate at a higher glycerol gradient percentage, between fractions 1 – 5, indicative of the subunit being a part of a larger complex, presumably the Integrator complex and/or conjugated to RNA Pol II (Figure 18). PPP2R1A showed two complexes forming, one at a higher glycerol percentage (fractions 3 and 4) and one at the middle point of the glycerol gradient (fractions 6 – 8). This could indicate that PPP2R1A was fractionated as a part of INTAC (the larger complex fraction), as well as part of free PP2A (the smaller complex fraction). NABP1 fractionated at a lower glycerol gradient percentage (fractions 10 – 12); this could indicate it being a part of the SOSS complex which is significantly smaller than Integrator. RNA Pol II, as expected, was observed to fractionate at a high glycerol percentage (fractions 1 – 4, especially fraction 3) suggestive of the whole RNA Pol II complex, possibly associated as well with Integrator (Figure 11). PPP2CA (fractions 6 and 7) was observed in the middle of the glycerol gradient percentages possibly indicating its complexing with the components of PP2A (Figure 11). NABP2, alike NABP1, fractionated at a lower glycerol gradient percentage (fractions 9 – 13); this could indicate it being a part of the SOSS complex which is significantly smaller than Integrator.

In comparison to the untreated samples, the four-hour treated samples showed distinct differences in signal for fractions collected (INTS1 and RNA Pol II) and additional bands at higher glycerol concentrations (PPP2R1A and PPP2CA). INTS1 showed increased signal for fractions 2 – 4, with an especially strong signal at fraction 2 for four hours treated samples compared to untreated (Figure 18). After four hours dTAG treatment, PPP2R1A did not have the higher molecular weight complex signal observed in the untreated sample (fractions 3 and 4) but did exhibit signal for the lower molecular weight complex (fractions 6 – 8) (Figure 18). If the higher molecular complex is, in fact, indicative of INTAC forming, it could suggest that INTS3 plays a part in recruiting

PPP2R1A to the complex. RNA Pol II showed similar signal for four hours treated, however, had a stronger signal for fraction 2 and displayed two discrete bands for fractions 3 and 4 which may be indicative of isoforms of RNA Pol II (Figure 18). PPP2CA showed an additional band in fraction 5 for treated compared to untreated and similar bands for fractions 6 and 7 (Figure 18). NABP1 and NABP2 showed similar signal at fractions 10 – 12 and 9 – 13, respectively (Figure 18).

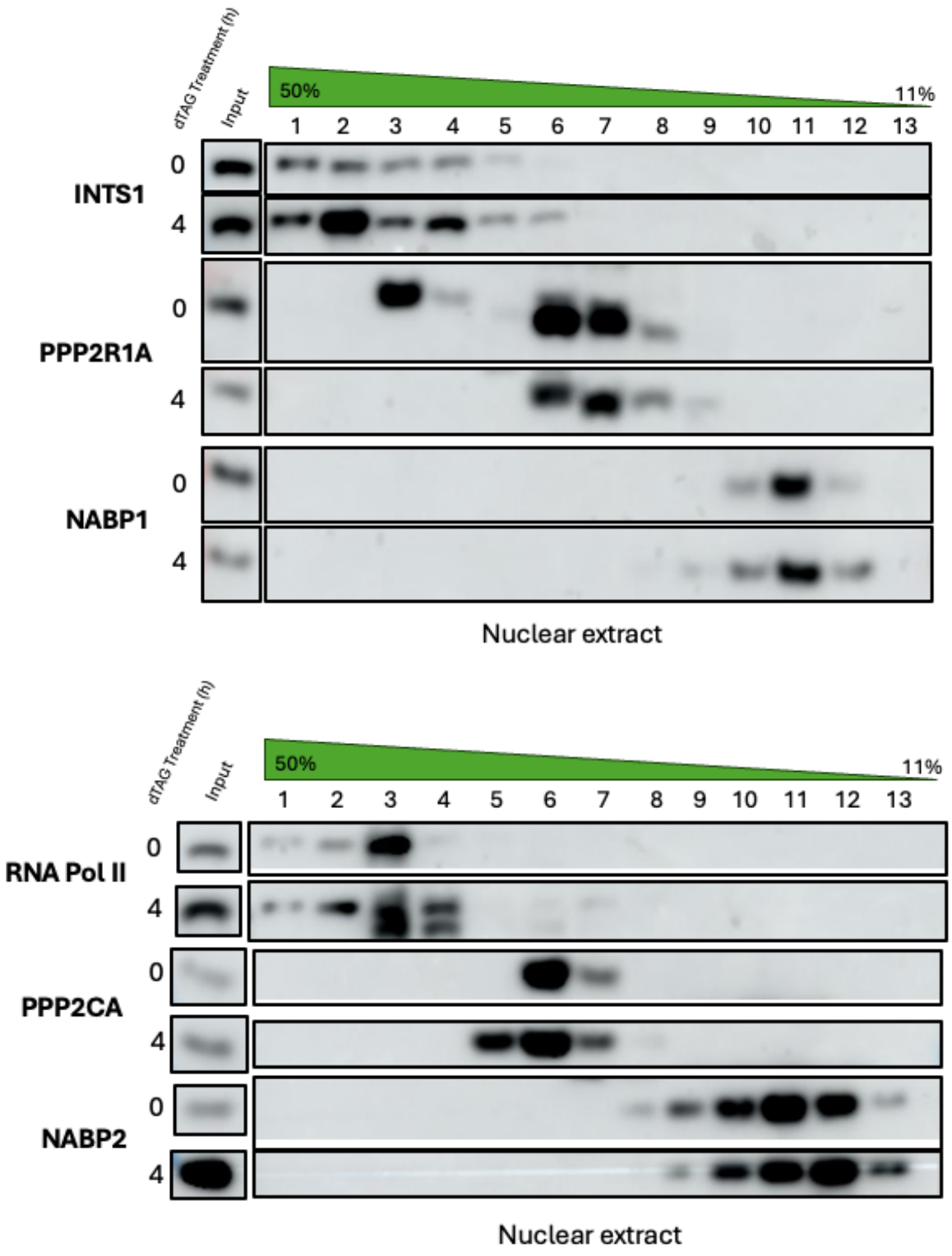



Figure 18: Glycerol gradient of nuclear extract for clone 10 cells untreated and after four hours dTAG treatment. For the untreated sample, INTS1 fractionated at a higher glycerol percentage (fractions 1 – 5) while for four hours treatment INTS1 showed an increase in signal for fractions 2 – 4, with an especially strong signal for fraction 2. At zero hours treatment PPP2R1A was

fractionated as two distinct complexes – one at a higher glycerol percentage between fractions 3 and 4, as well as between fractions 6 – 8 while at four hours treatment only the complex between fractions 6 – 8 was observed. NABP1 showed similar signal for both untreated and treated samples, having signal at a low glycerol percentage between fractions 10 – 12. RNA Pol II was observed to complex at a high glycerol percentage for both untreated and treated sample (between fractions 1 – 4) but showed more intense signal at fraction 2 for the treated sample, along with two distinct bands forming for fractions 3 – 4. PPP2CA was observed to complex in the middle of the glycerol gradient (fractions 5 – 7), with an additional band seen for fraction 5. NABP2 showed similar signal for both untreated and treated samples, having signal at a low glycerol percentage between fractions 9 – 13.

**INTS3 ablation is correlated with changes to the transcriptome, especially with respect to immediate-early response genes (IERGs)**

In order to assess the effect of INTS3 ablation on the transcriptome, Quant-seq was performed. Quant-seq is a high-throughput RNA sequencing technique tailored to generate sequences near the 3`-end of polyadenylated RNA, making it highly effective for quantifying gene expression and detecting transcriptomic changes. The high effectiveness of Quant-seq in terms of gene expression quantification is a result of its accurate counting of mRNA molecules, allowing for precise measurement of gene expression levels. This method is particularly useful for differential expression analysis, identifying genes that are up-or down-regulated in response to specific treatments, conditions, and stimuli. Additionally, Quant-seq provides information on transcript abundance, highlighting which genes are highly or lowly expressed in a sample. The technique also offers insights into alternative polyadenylation sites, which can influence mRNA stability, localization, and translation efficiency, thereby affecting gene expression regulation.

Quant-seq was performed for clone 10 cells untreated with dTAG, and for those treated with dTAG for 30 minutes, one hour and four hours. Clone 10 cells were seeded at the same density and treated with the same stock of $dTAG^{V}$-1 molecule. All samples were processed together under the same conditions according to the protocol. Data obtained from Illumina sequencing were analyzed with DESeq2. PCA plotting shows the samples clustering according to treatment time, reinforcing the reliability of the data. Specifically, the samples for zero hours showed clustering, those treated for 30 minutes and one hour were observed to cluster, and samples for four hours treatment

additionally clustered, although more weakly in comparison to the other conditions with an outlier for the second replicate for four-hours treated (Figure 19).

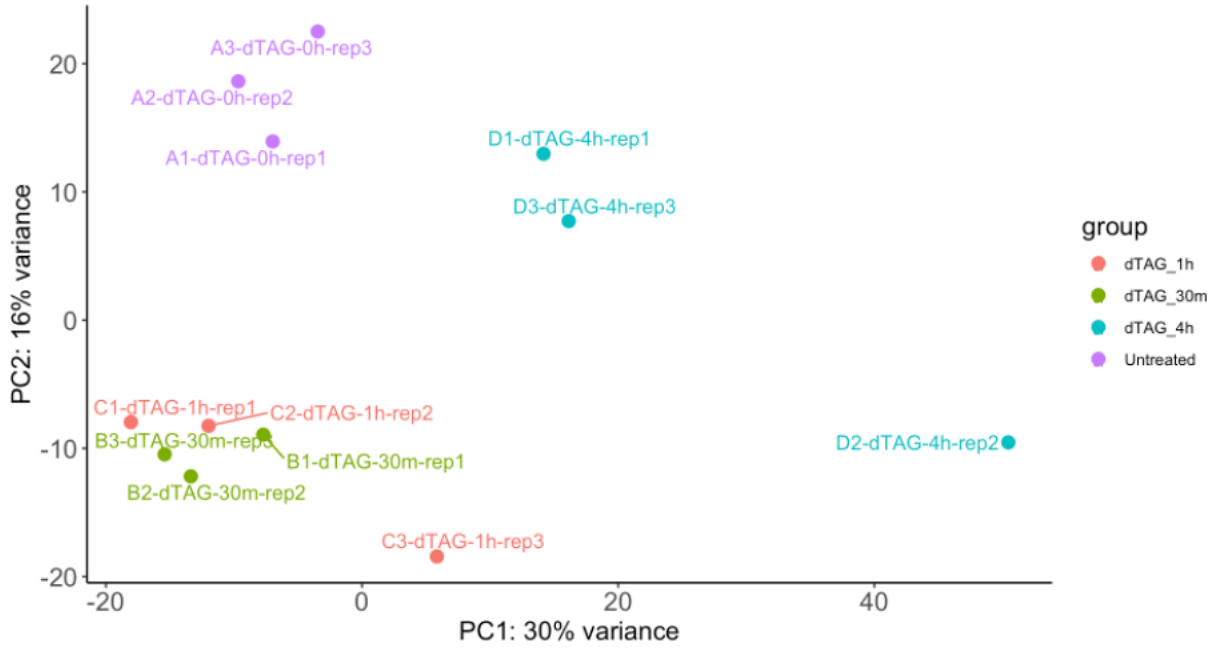



Figure 19: PCA plot of clone 10 cells treated with dTAG. Replicates for zero hours (untreated) cluster (purple), alongside replicates for 30 minutes (green) and one-hour (red) dTAG treatment which cluster distinctly together. Two replicates for dTAG treated for four hours (blue) cluster together with one outlier (dTAG 4h replicate 2).

Differential gene expression analysis for comparing 30 minutes, one hour and four hours dTAG treatment against untreated samples yielded a differential gene expression profile showing a subset of genes up- and down-regulated for each condition. Volcano plots show that for 30 minutes dTAG treatment 84 genes were significantly upregulated, and 34 genes were significantly downregulated; for one-hour dTAG treatment 75 genes were significantly upregulated, and 28 genes were significantly downregulated; and for four hours dTAG treatment 96 genes were significantly upregulated, and 47 genes were significantly downregulated (Figure 20). The subset of most upregulated genes for 30 minutes and one-hour dTAG treated samples included EGR1, FOS, JUND, FOSB, JUNB, EGR2, ATF3, NR4A1, BBC3 and RNU5B-1 (Figure 20). The subset of most downregulated genes included BIRC3, CCL2, and STC1 for those shared between dTAG treated for 30 minutes overlapping with one hour (Figure 20). For four hours treatment, upregulated genes of interest included INTS5, RNU5B-1 and FOSB; over four hours the cohort of

immediate early response genes was not observed to be upregulated as for 30 minutes and one hour. In particular, EGR1 was observed to be strongly upregulated after one-hour dTAG treatment, and more moderately so after 30 minutes dTAG treatment (Figure 20).

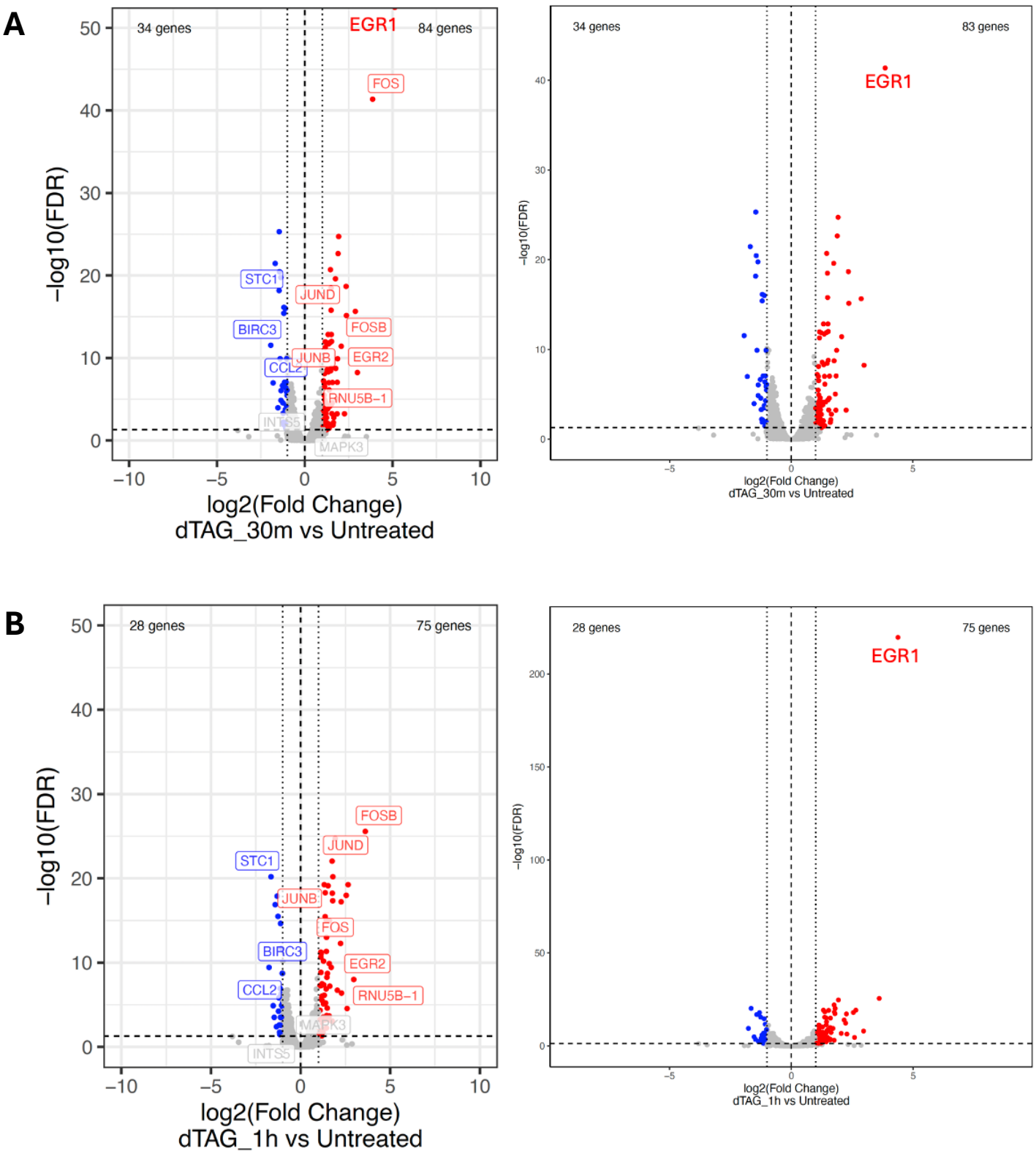



Figure 20 (continued on following page)

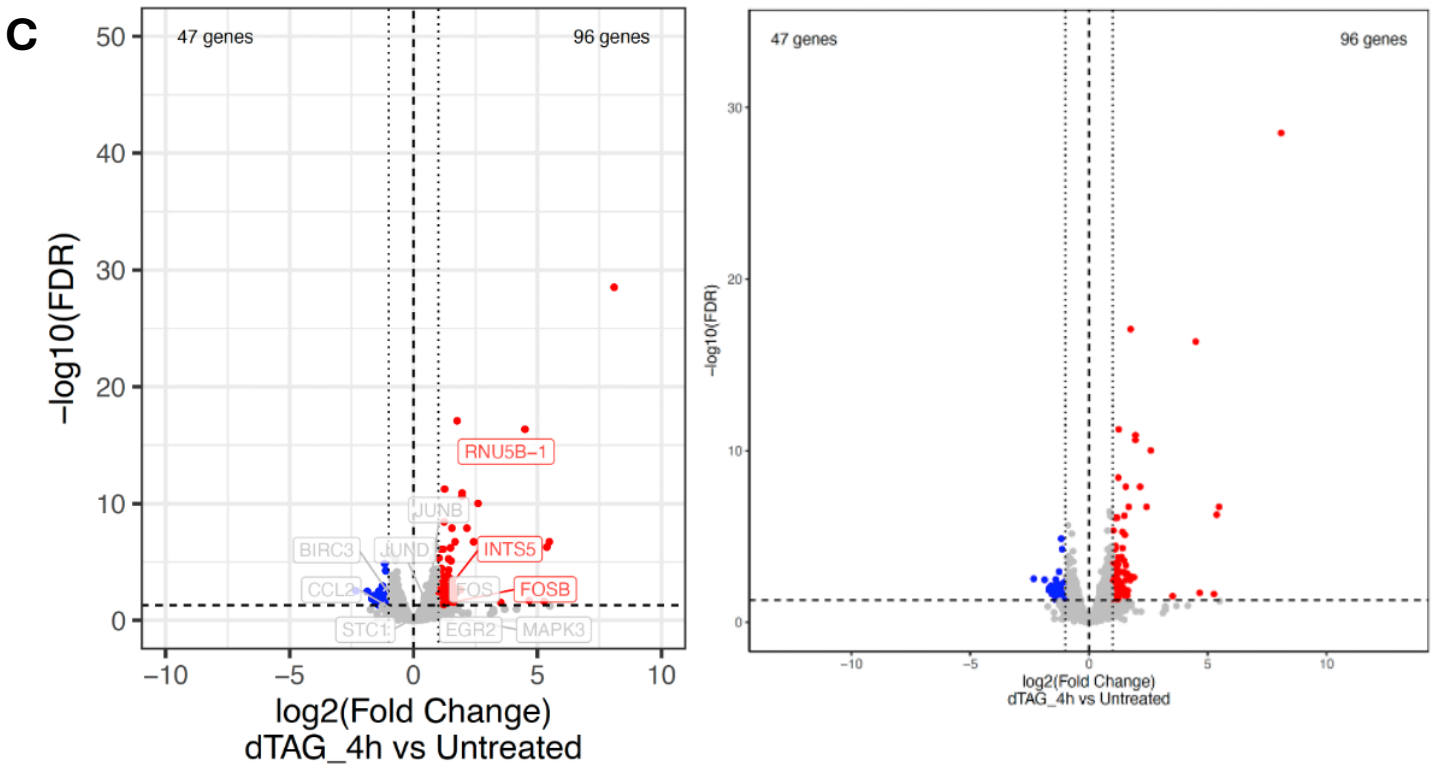



Figure 20: Volcano plots differentially expressed genes for 30 minutes, 1-hour and 4-hour dTAG treatment vs. untreated. (A) After 30 minutes dTAG treatment 84 genes were significantly upregulated, and 34 genes were significantly downregulated. (B) After one-hour dTAG treatment 75 genes were significantly upregulated, and 28 genes were significantly downregulated. Of the upregulated genes, EGR1 was most significantly upregulated alongside various immediate-early response genes such as FOSB, JUND, JUNB, FOS, and EGR2. (C) After 4 hours dTAG treatment 96 genes were significantly upregulated, and 47 genes were significantly downregulated. The immediate early response genes upregulated in (A) and (B) were not observed for four hours dTAG treatment.

In order to assess whether treatment with the dTAG molecule itself was eliciting an upregulated reaction for immediate early response genes, wild-type OVCAR8 cells were treated with dTAG identically to the experiment above, with treatment times of zero hours (untreated), 30 minutes, one hour and four hours to act as a negative control. wtOVCAR8 cells were seeded at the same density and treated with the same stock of dTAG$^{V}$-1 molecule. All samples were processed together under the same conditions according to the protocol. Data obtained from Illumina sequencing were analyzed with DESeq2. PCA plotting shows the samples clustering according to treatment time, reinforcing the reliability of the data. Specifically, the samples for zero hours showed clustering, as well as those treated for 30 minutes, one hour and four hours (Figure 21). Volcano plots show that, as expected, for 30 minutes dTAG treatment 0 genes were significantly

upregulated or significantly downregulated, and for one-hour dTAG treatment, 3 genes were significantly upregulated, and 9 genes were significantly downregulated. However, for four hours dTAG treatment 175 genes were significantly upregulated, and 319 genes were significantly downregulated (Figure 22) which was unexpected as a result. This may due to a cellular response to the dTAG molecule for the long duration of treatment, off-target effects occurring, or the induction of cellular stress causing the increase in differentially expressed genes .

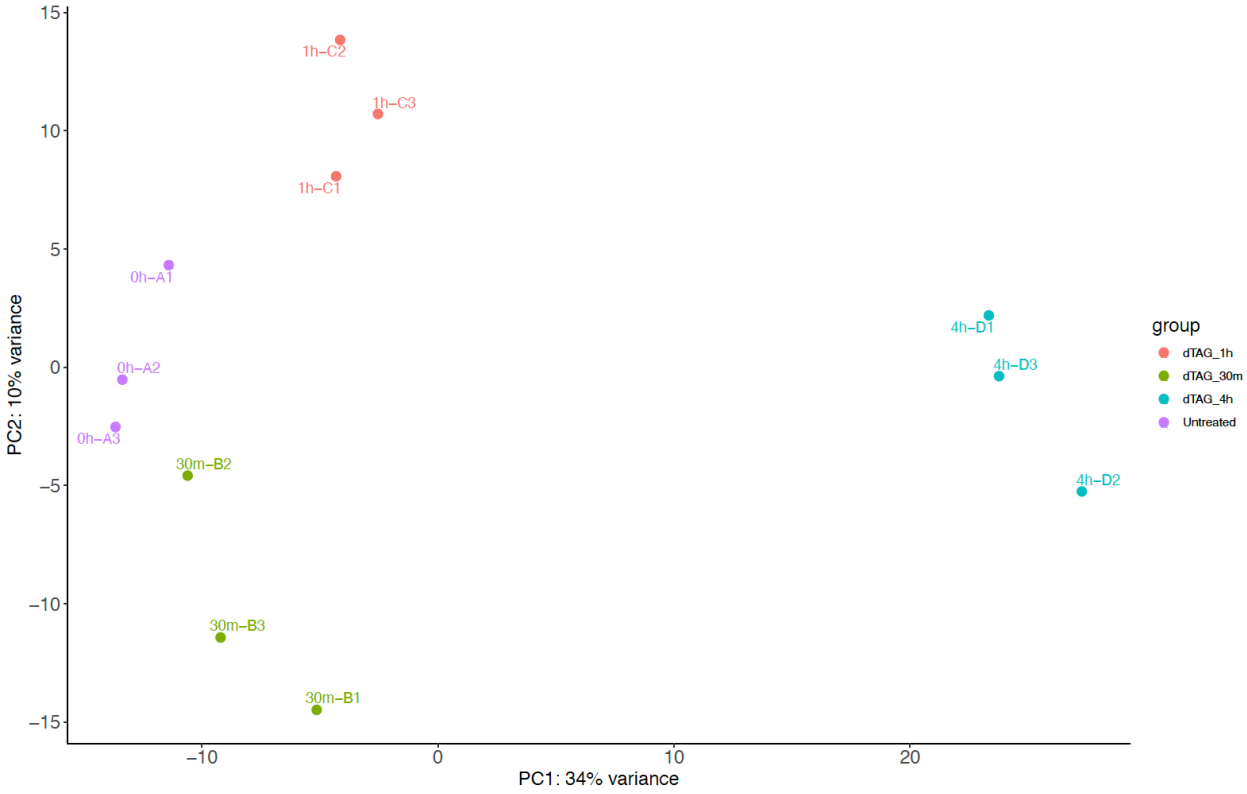



Figure 21: PCA plot of wtOVCAR8 cells treated with dTAG. Replicates for zero hours (untreated) cluster (purple), alongside replicates for 30 minutes (green), one hour (red) and four hours (purple).

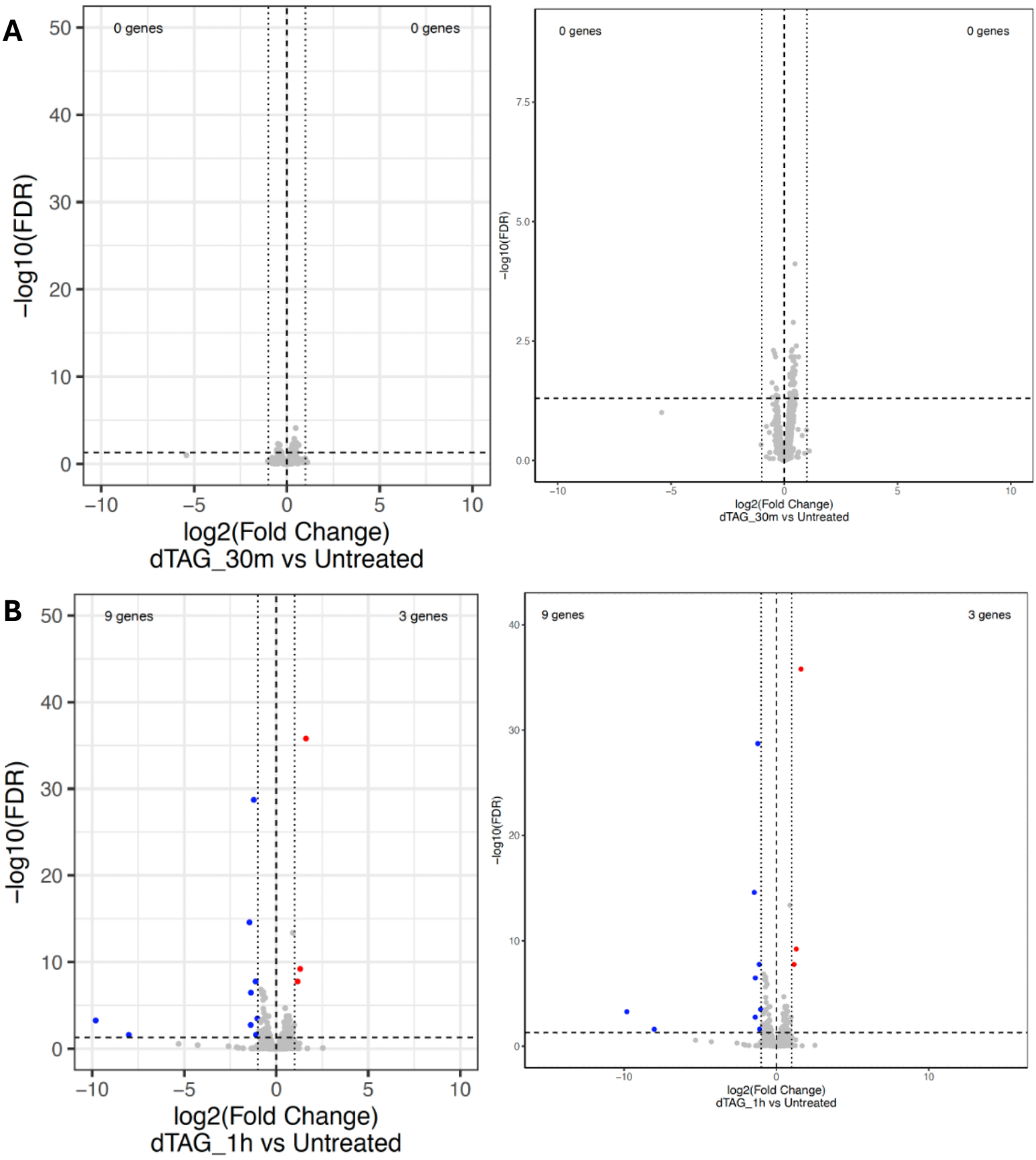



Figure 22 (continued on following page)

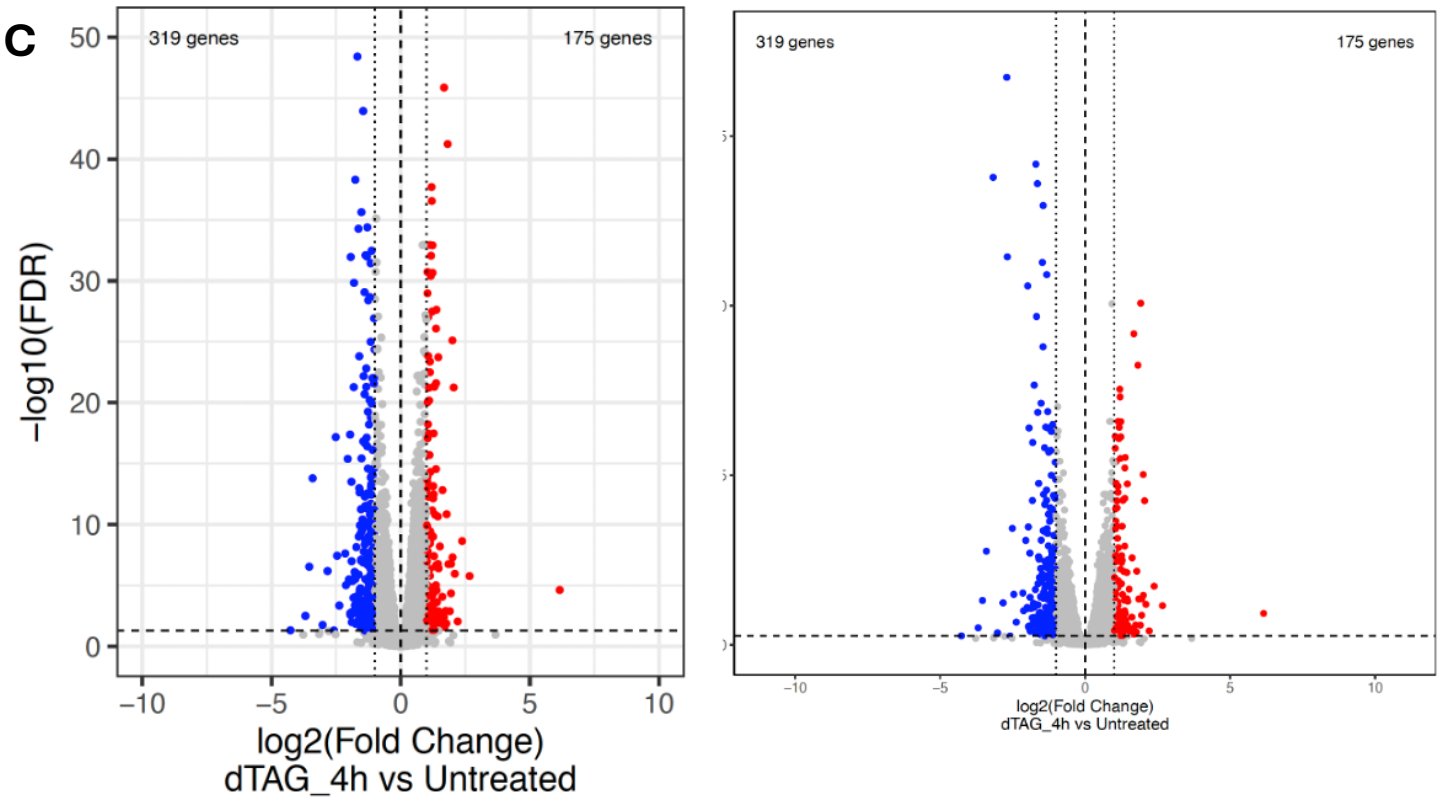



Figure 22: Volcano plots differentially expressed genes for 30 minutes, 1-hour and 4-hour dTAG treatment vs. untreated in wtOVCAR8. (A) After 30 minutes dTAG treatment 0 genes were significantly upregulated or downregulated. (B) After one-hour dTAG treatment 3 genes were significantly upregulated, and 9 genes were significantly downregulated. Of the upregulated genes, EGR1 or other immediate early response genes were not observed to be up- or down-regulated. (C) After 4 hours dTAG treatment 175 genes were significantly upregulated, and 319 genes were significantly downregulated.

Gene ontology (GO) analysis was performed for the upregulated gene sets for 30 minutes and one-hour dTAG treatment against untreated samples in clone 10 cells. The biological processes with the lowest p-values identified by GO were all related to regulation of transcription. The upregulated gene set for 30 minutes vs. untreated resolved the following biological processes: regulation of transcription by RNA polymerase II ($p = 1.386 \times 10^{-9}$), regulation of DNA-templated transcription ($p = 1.965 \times 10^{-9}$), positive regulation of transcription by RNA polymerase II, ($p = 5.210 \times 10^{-8}$), negative regulation of transcription by RNA polymerase II ($p = 5.872 \times 10^{-7}$), positive regulation of DNA-templated transcription ($p = 7.121 \times 10^{-7}$), and negative regulation of DNA-templated transcription ($p = 1.007 \times 10^{-6}$) (Figure 23). The upregulated gene set for one hour vs. untreated resolved the following biological processes: regulation of transcription by RNA polymerase II

($p = 1.007 \times 10^{-11}$), negative regulation of transcription by RNA polymerase II ($p = 3.442 \times 10^{-11}$), regulation of DNA-templated transcription ($p = 9.554 \times 10^{-11}$), negative regulation of DNA-templated transcription ($p = 7.105 \times 10^{-10}$), positive regulation of transcription by RNA polymerase II ($p = 7.806 \times 10^{-9}$), and positive regulation of DNA-templated transcription ($p = 1.908 \times 10^{-8}$) (Figure 23).

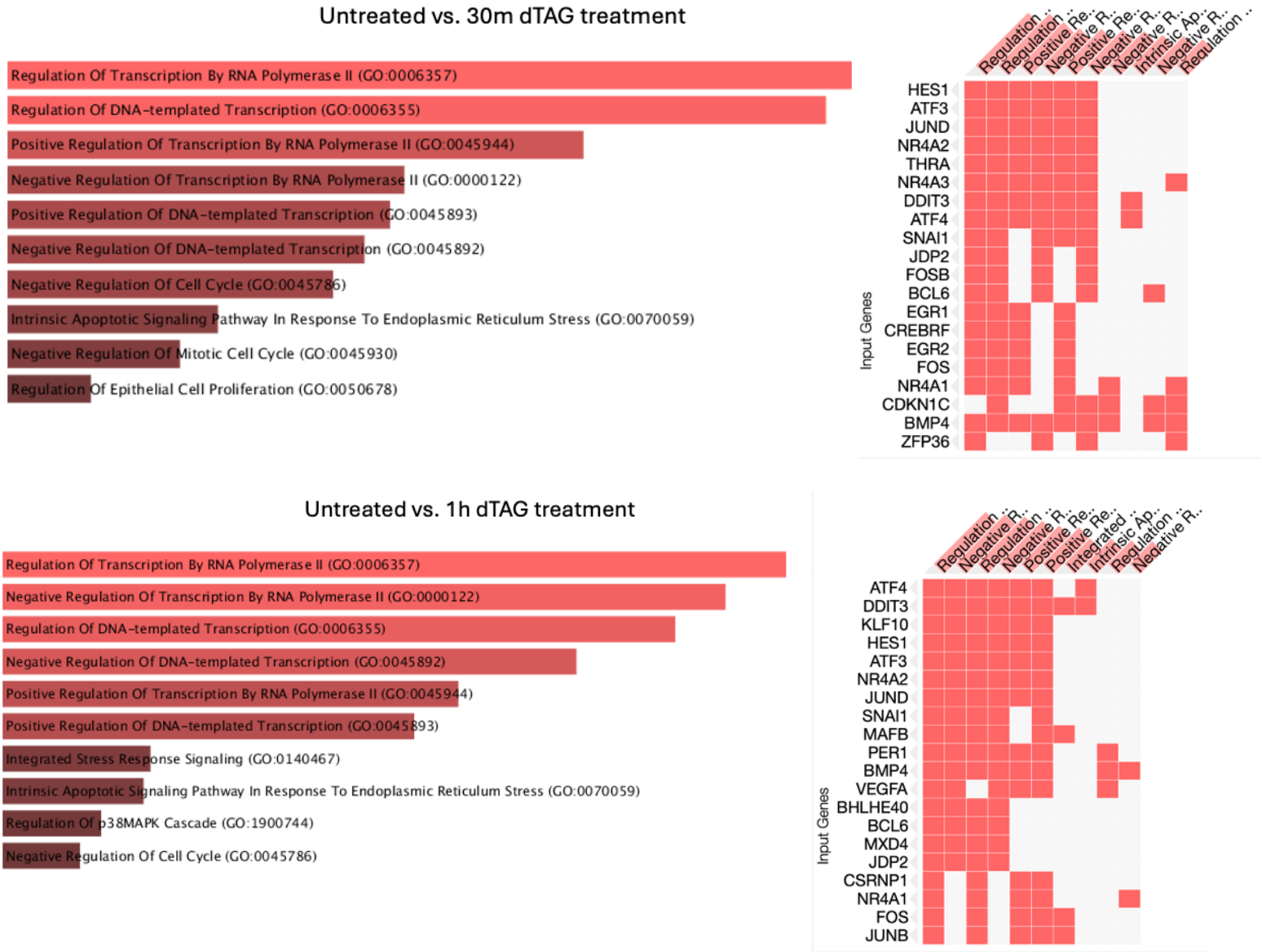



Figure 23: Gene ontology analysis for upregulated gene for 30 minutes and one-hour dTAG treated cells vs. untreated in clone 10 cells. For both 30 minutes and one-hour dTAG treated samples vs. untreated, the GO biological processes included those involved in transcriptional regulation such as: regulation of transcription by RNA polymerase II, regulation of DNA-templated transcription, positive regulation of transcription by RNA polymerase II, negative regulation of transcription by RNA polymerase II, positive regulation of DNA-templated transcription, and negative regulation of DNA-templated transcription

**Splicing factor 3 subunits are differentially pulled down dependent on INTS3 ablation**

As it was observed that various Integrator subunits were upregulated after dTAG treatment, and that complex formation was affected by INTS3 ablation, IP followed by mass spectrometry (IP-MS) was conducted to probe whether complex composition of Integrator or RNAPII was affected by INTS3 ablation. In addition, IP-MS was conducted to observe if there were any additional interacting partners of INTS3 that were gained or lost as a result of INTS3 ablation. INTS3 was completely degraded and was only pulled down in the untreated sample (Figure 24). Results show that the Integrator and RNAPII complexes are largely unaffected by INTS3 ablation, although Integrator subunit 14 (INTS14) shows slightly more abundance for the dTAG treated sample. Furthermore, most PP2A subunits had similar abundances pulled down in both the treated and untreated samples, except for PPP2R3A, a B-subunit of the PP2A complex and not a part of INTAC. NABP2 was observed to only be pulled down in the untreated sample, as expected since INTS3 is a scaffold for the protein in the SOSS complex (Figure 24). Of interest were splicing factor 3 subunits which showed varying abundances dependent on dTAG treatment. In particular, SF3B3 and SF3B6 were more highly abundant in the untreated sample, with a moderate difference seen for SF3A3, SF3A1, and SF3B1. SF3B2 was more highly abundant in the treated sample. These splicing factor 3 subunits are pertinent as it may suggest a role of INTS3 in U snRNA biogenesis, a process already suspected in that INTS3 may play a role.

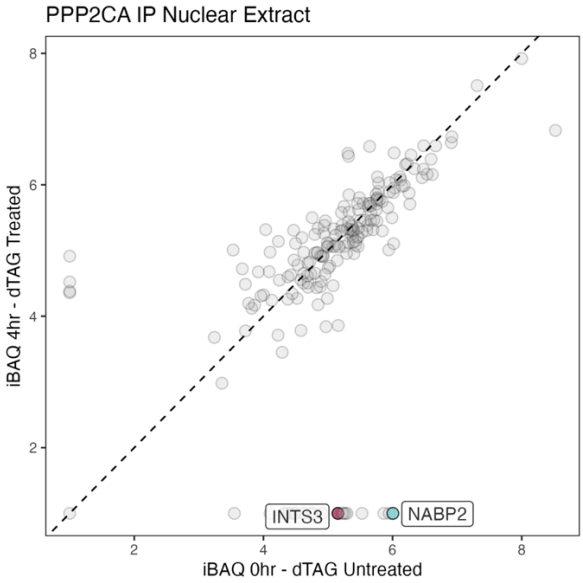



Figure 24 (continued on following page)

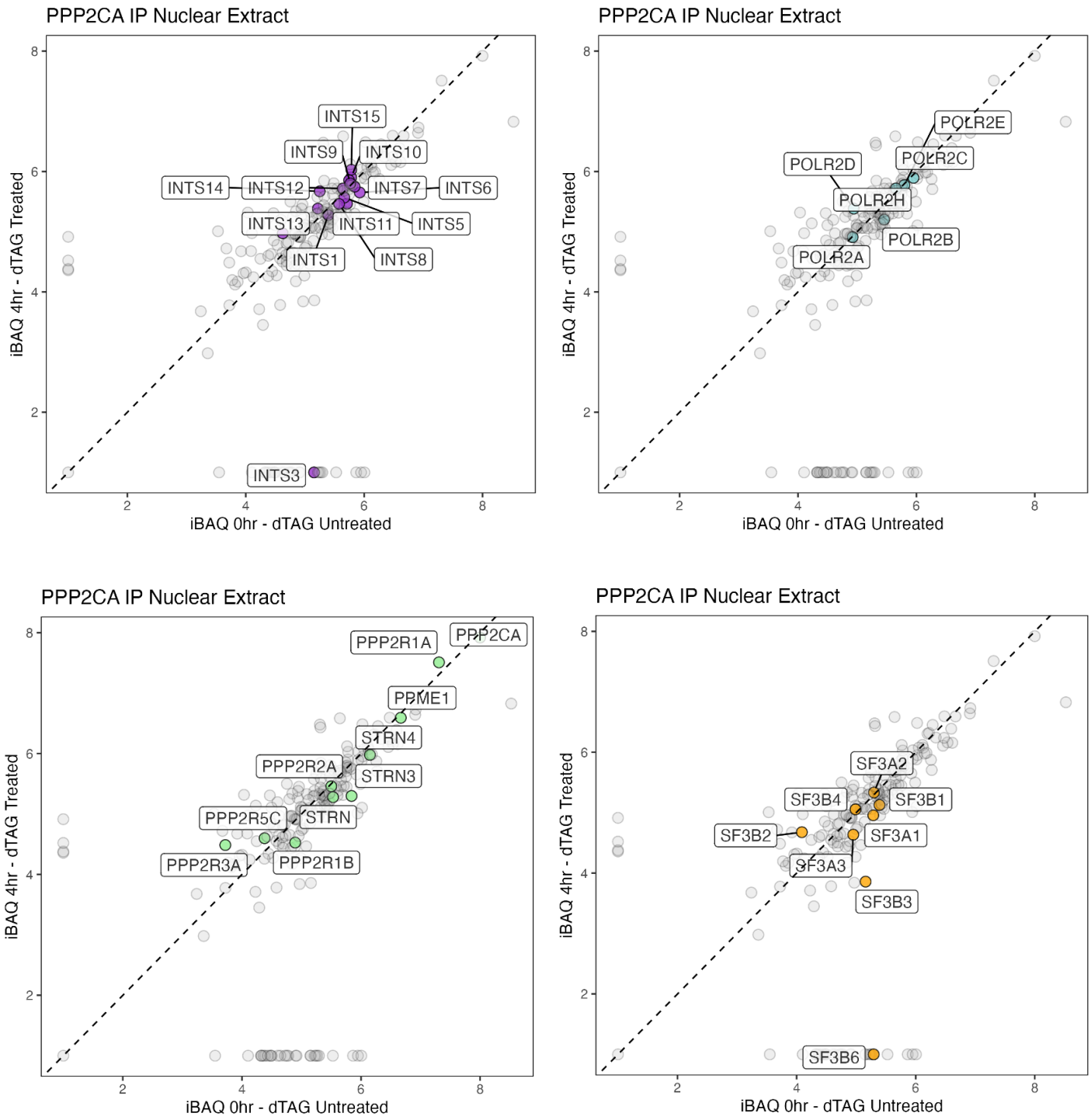



Figure 24: Immunoprecipitation followed by mass spectrometry for untreated against four hours dTAG treated samples. INTS3 and NABP2 were only pulled down in the untreated sample showing INTS3 ablation was complete. Of the Integrator subunits, the complex (purple) as a whole was unaffected by dTAG treatment and subsequent INTS3 ablation, except for INTS14 which showed minorly more abundance for the dTAG treated sample. Additionally, RNAPII (blue) was largely unaffected by INTS3 ablation. PP2A subunits (green) had similar abundances

pulled down in both the treated and untreated samples, except for PPP2R3A, a B-subunit of the PP2A complex. Splicing factor 3 A and B (SF3A and SF3B) (orange) subunits showed varying abundances depending on treatment. SF3B3 and SF3B6 were more highly abundant in the untreated sample, with a moderate difference seen for SF3A3, SF3A1, and SF3B1. SF3B2 was more highly abundant in the treated sample.

**The SOSS complex is present alongside INTS3 across the genome, found predominantly at a motif which controls IERG transcription**

In order to observe if the SOSS complex components were present alongside RNA Pol II and Integrator across the genome, ChIP-seq was performed. ChIP-seq, or Chromatin Immunoprecipitation coupled with sequencing, is a technique that allows for the investigation of protein-DNA interactions on a genome-wide scale. By combining chromatin immunoprecipitation with next-generation sequencing (NGS), ChIP-seq is a tool allowing for the study of transcription factor binding sites, histone modifications, and other chromatin associated proteins.

ChIP-seq was performed with antibodies specific to RNA Pol II (RBP1), INTS3, and NABP1. For RNA Pol II, INTS3 and NABP1, peaks were enriched around the TSS of the gene as observed in the representative gene tracks of FOS, EGR1 and JUN and across the most active genes in OVCAR8 cells (Figures 25 and 26). Furthermore, RNA Pol II, INTS3 and NABP1 are co-localized at the TSS of the most active genes in OVCAR8 cells (Figures 25 and 26). This indicates that the SOSS complex is present at the promoter region of active genes and may be a part of Integrator at such promoter sites.

In addition, it was observed that INTS3 and NABP1 were found to be present in various proportions across promoter and enhancer regions across the genome. INTS3 was found to be present at roughly an equal proportion of promoters and enhancers (48 % promoters vs. 52 % enhancers) while NABP1 was observed to be present at a majority of enhancers as compared to promoters (17% promoters vs. 83 % enhancers) (Figure 27). Despite NABP1 being localized at a majority of enhancers, there was significant overlap of 5163 enhancer regions where INTS3 and NABP1 were observed to overlap (Figure 27).

Furthermore, HOMER was employed for motif discovery and *de novo* motif discovery to probe at which motifs INTS3 and NABP1 were most commonly intersecting. It was found by *de novo* motif discovery that the AP-1 motif (TGA(G/C)TCA) was most highly enriched as a target for INTS3 and NABP1 binding, alongside an unknown motif, ZFP809 and HIF1α (Figure 28). This was bolstered by known motif discovery that showed INTS3 and NABP1 associating at motifs of the same TGA(G/C)TCA sequence which matches IERGs motifs such as ATF3, FOSL2, JUN, AP-1, and JUNB (Figure 28). This specific motif is also indicative of active transcription which may suggest INTS3 and NABP1 are involved in the process.

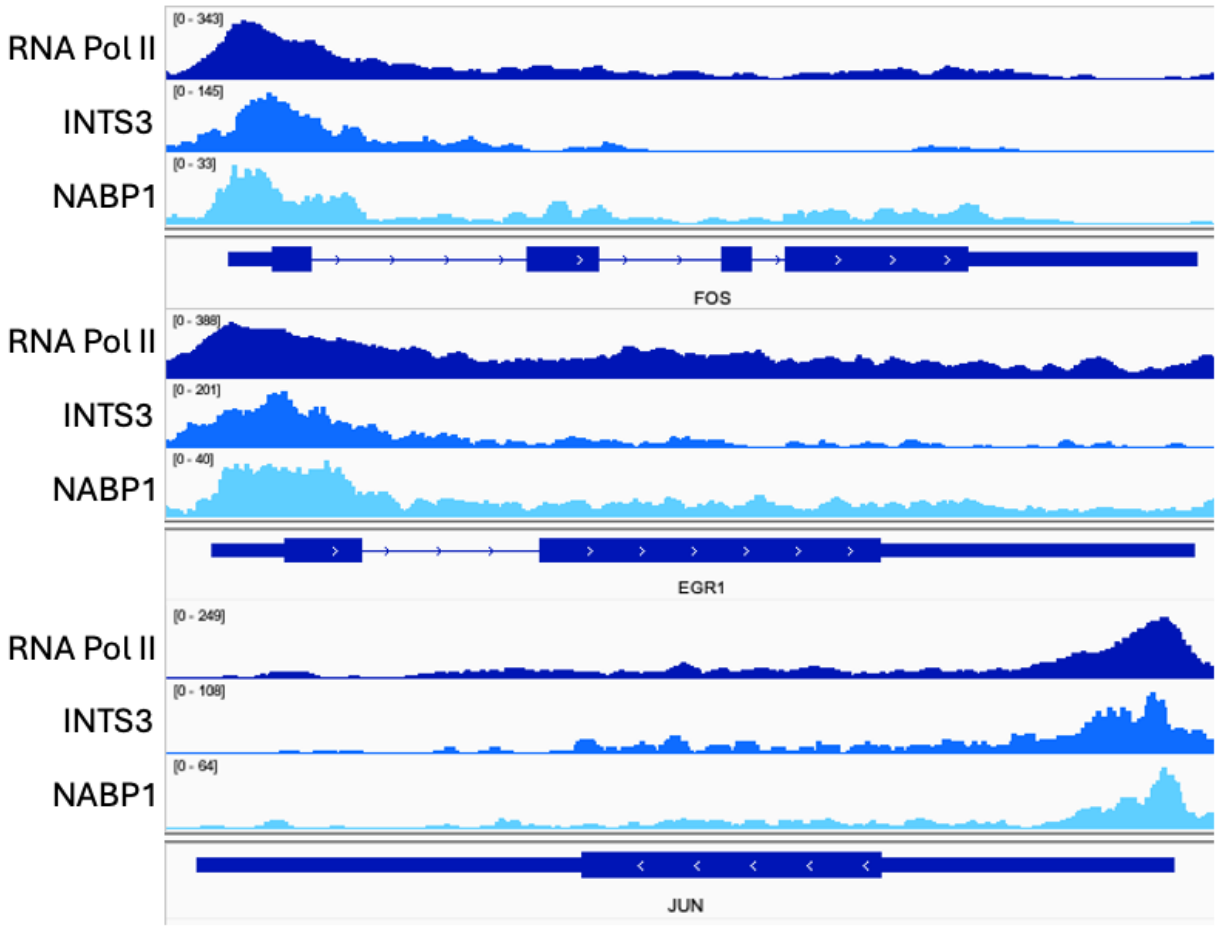



Figure 25: Representative gene tracks for RNA Pol II, INTS3 and NABP1 binding across the genome. For RNA Pol II (dark blue), INTS3 (blue) and NABP1 (light blue), peaks were enriched around the TSS of the gene as observed in the representative gene tracks of FOS, EGR1 and JUN

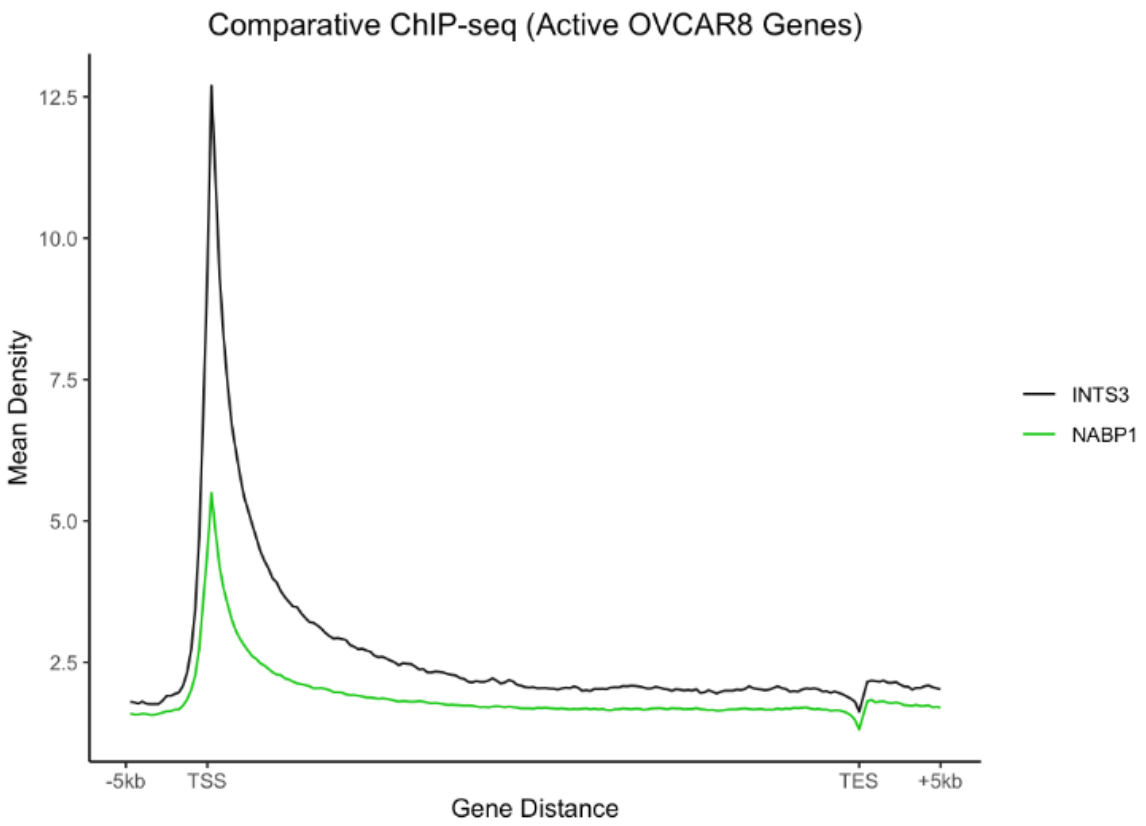
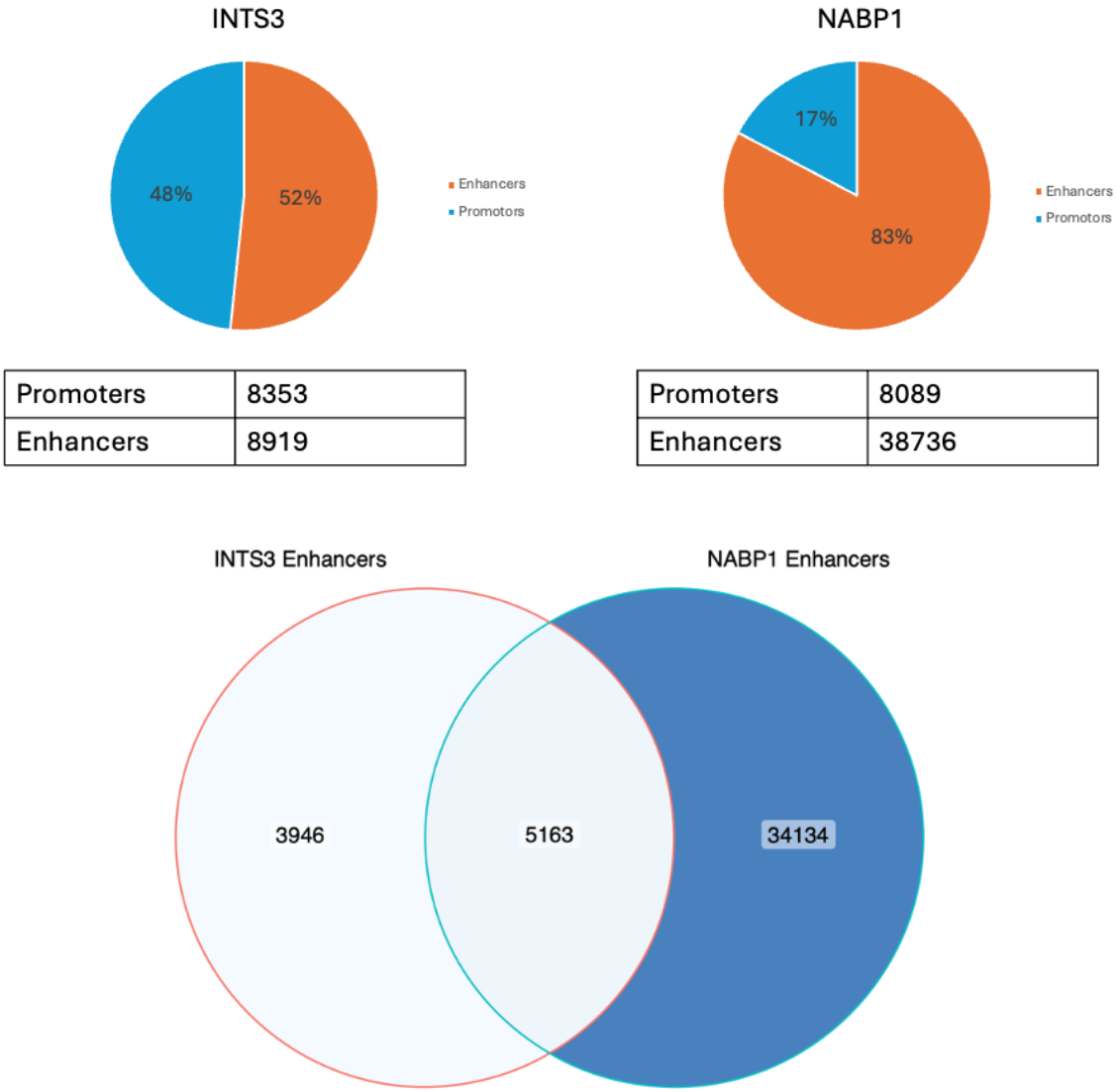



Figure 26: Average profile of INTS3 and NABP1 binding across active genes in OVCAR8. INTS3 (black) and NABP1 (green) co-localize at the TSS of active genes in OVCAR8. INTS3 and NABP1 are observed to be present at lower proportions in the gene body or at the TES.




| | |
|---|---|
| Promoters | 8353 |
| Enhancers | 8919 |

| | |
|---|---|
| Promoters | 8089 |
| Enhancers | 38736 |

Figure 27: Promoter and enhancer binding proportions for INTS3 and NABP1. INTS3 was found to be present at roughly an equal proportion of promoters and enhancers (48 % promoters vs. 52 % enhancers) while NABP1 was observed to be present at a majority of enhancers as compared to promoters (17% promoters vs. 83 % enhancers). Despite NABP1 being localized at a majority of enhancers, there was significant overlap of 5163 enhancer regions where INTS3 and NABP1 were observed to overlap.

**A**

| Rank | Motif | P-value | log P-pvalue | % of Targets | % of Background | STD(Bg STD) | |
|---|---|---|---|---|---|---|---|
| 1 |  | 1e-15 | -3.649e+01 | 5.23% | 2.68% | 53.9bp (61.5bp) | AP-1 |
| 2 |  | 1e-13 | -3.036e+01 | 0.29% | 0.01% | 62.9bp (35.4bp) | No known or similar motifs |
| 3 |  | 1e-12 | -2.846e+01 | 8.07% | 5.14% | 54.5bp (62.8bp) | ZFP809 |
| 4 |  | 1e-12 | -2.792e+01 | 6.95% | 4.28% | 51.6bp (61.0bp) | HIF1α |

**B**

| Rank | Motif | Name | P-value | log P-pvalue | q-value (Benjamini) | # Target Sequences with Motif | % of Targets Sequences with Motif |
|---|---|---|---|---|---|---|---|
| 1 |  | Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer | 1e-13 | -3.078e+01 | 0.0000 | 182.0 | 5.32% |
| 2 |  | Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer | 1e-12 | -2.839e+01 | 0.0000 | 209.0 | 6.11% |
| 3 |  | Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer | 1e-12 | -2.770e+01 | 0.0000 | 169.0 | 4.94% |
| 4 |  | Fosl2(bZIP)/3T3L1-Fosl2-ChIP-Seq(GSE56872)/Homer | 1e-11 | -2.719e+01 | 0.0000 | 135.0 | 3.95% |
| 5 |  | Bach2(bZIP)/OCILy7-Bach2-ChIP-Seq(GSE44420)/Homer | 1e-9 | -2.250e+01 | 0.0000 | 95.0 | 2.78% |
| 6 |  | Jun-AP1(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer | 1e-9 | -2.240e+01 | 0.0000 | 100.0 | 2.92% |
| 7 |  | AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer | 1e-8 | -2.045e+01 | 0.0000 | 220.0 | 6.43% |
| 8 |  | JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer | 1e-8 | -2.036e+01 | 0.0000 | 172.0 | 5.03% |
| 9 |  | BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer | 1e-8 | -1.952e+01 | 0.0000 | 186.0 | 5.44% |

Figure 28: *De novo* and known motif discovery for intersecting motifs of INTS3 and NABP1. (A) *De novo* motif discovery found that the AP-1 motif (TGA(G/C)TCA) was most highly enriched as a target for INTS3 and NABP1 binding, alongside an unknown motif, ZFP809 and HIF1α. (B) INTS3 and NABP1 associated at motifs of the same TGA(G/C)TCA sequence which matches IERGs motifs such as ATF3, FOSL2, JUN, AP-1, and JUNB.

**INTS3 ablation affects RNAPII occupancy and increases pSer2 and pSer5 states across the genome**

In order to assess the effects of INTS3 ablation on RNAPII dynamics, ChIP-seq was performed on clone 10 cells in which RNAPII, pSer2 and pSer5 signal was monitored over a short time course of 0, 30 and 60 minutes. Average profile plots of RNAPII occupancy (Figure 29), and pSer2 (Figure 30) and pSer5 (Figure 31) phosphorylation levels, over the most active 5000 genes across the genome were generated.

INTS3 degradation affected RNAPII occupancy and pSer2 and pSer5 phosphorylation levels. The RNAPII profile shows a pronounced enrichment at the TSS at 0 hours, indicative of promoter-proximal pausing, a key regulatory step in transcription initiation (Figure 29). However, after 30 minutes and 60 minutes of treatment, there is a notable reduction in RNAPII density both at the TSS and across the gene body, suggesting that INTS3 ablation impairs RNAPII's ability to transition from initiation to productive elongation (Figure 29). The decline in RNAPII levels at the TES further indicates that transcriptional termination may also be affected, potentially due to a reduced transcriptional processivity or overall disruption in RNAPII function.

In contrast, the pSer2 and pSer5 profiles—which marks RNAPII's phosphorylation at serine 2 and 5, modifications associated with elongation and initiation—exhibit a different trend (Figures 30 and 31). While the pSer2 signal is present at a baseline level at 0 hours, it significantly increases after 30 and 60 minutes of INTS3 degradation. This increase is observed across the TSS, gene body, and TES, indicating that despite the overall reduction in RNAPII occupancy, RNAPII remains more heavily phosphorylated at serine 2, suggesting an enhanced elongation phase (Figure 23). In addition, pSer5 is present at baseline at 0 hours, also displaying an increase after 30 and 60 minutes. This increase is further observed across the TSS, gene body and TES, indicating an enhanced initiation and elongation phase. The increase in pSer5 levels at 30 minutes and 60 minutes compared to 0 hours, suggests that INTS3 ablation enhances both transcription initiation and early elongation phases (Figure 31).

The combined data from the RNAPII, pSer2 and pSer5 average profiles suggest a complex response to INTS3 ablation. The decrease in overall RNAPII occupancy indicates that fewer RNAPII molecules are engaged in transcription, potentially due to impaired initiation or premature termination. However, the concurrent increase in pSer2 and pSer5 levels implies that the remaining RNAPII molecules are more transcriptionally active, as indicated by enhanced Ser2 and Ser5 phosphorylation, which is crucial for initiation and elongation.

Furthermore, a traveling ratio for RNAPII at 0h, 30 minutes and 60 minutes dTAG treatment was generated (Figure 32). It was observed that for INTS3 ablation for 30 and 60 minutes the traveling was reduced (Figure 32). The reduction in the traveling ratio of RNAPII observed at 30 minutes and 60 minutes indicates a compromised transcriptional elongation process under INTS3 ablation.

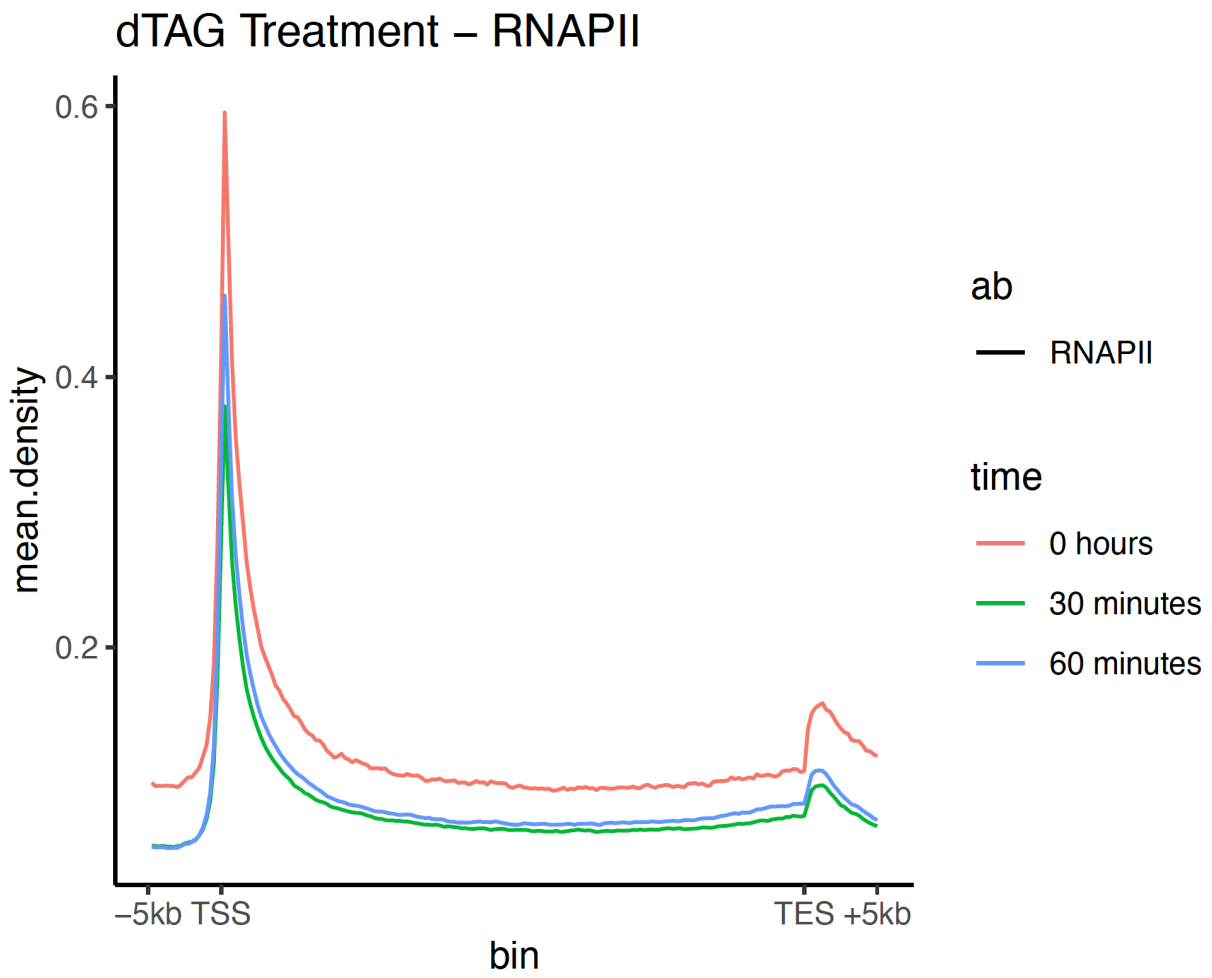



Figure 29: Average profile of RNAPII occupancy across the transcription start site (TSS), gene body, and transcription end site (TES) in clone 10 cells. The graph shows the mean density of RNA Polymerase II (RNAPII) at three time points: 0 hours, 30 minutes, and 60 minutes post-treatment. The x-axis represents the distance from the TSS, spanning from -5 kb upstream of the TSS to +5 kb downstream of the TES. The y-axis shows the normalized mean RNAPII density. At 0 hours, RNAPII exhibits strong enrichment at the TSS, indicative of promoter-proximal pausing. Following INTS3 ablation, there is a noticeable decrease in RNAPII occupancy at both the TSS and across the gene body at 30 minutes and 60 minutes, suggesting a reduction in RNAPII recruitment and/or processivity over time.

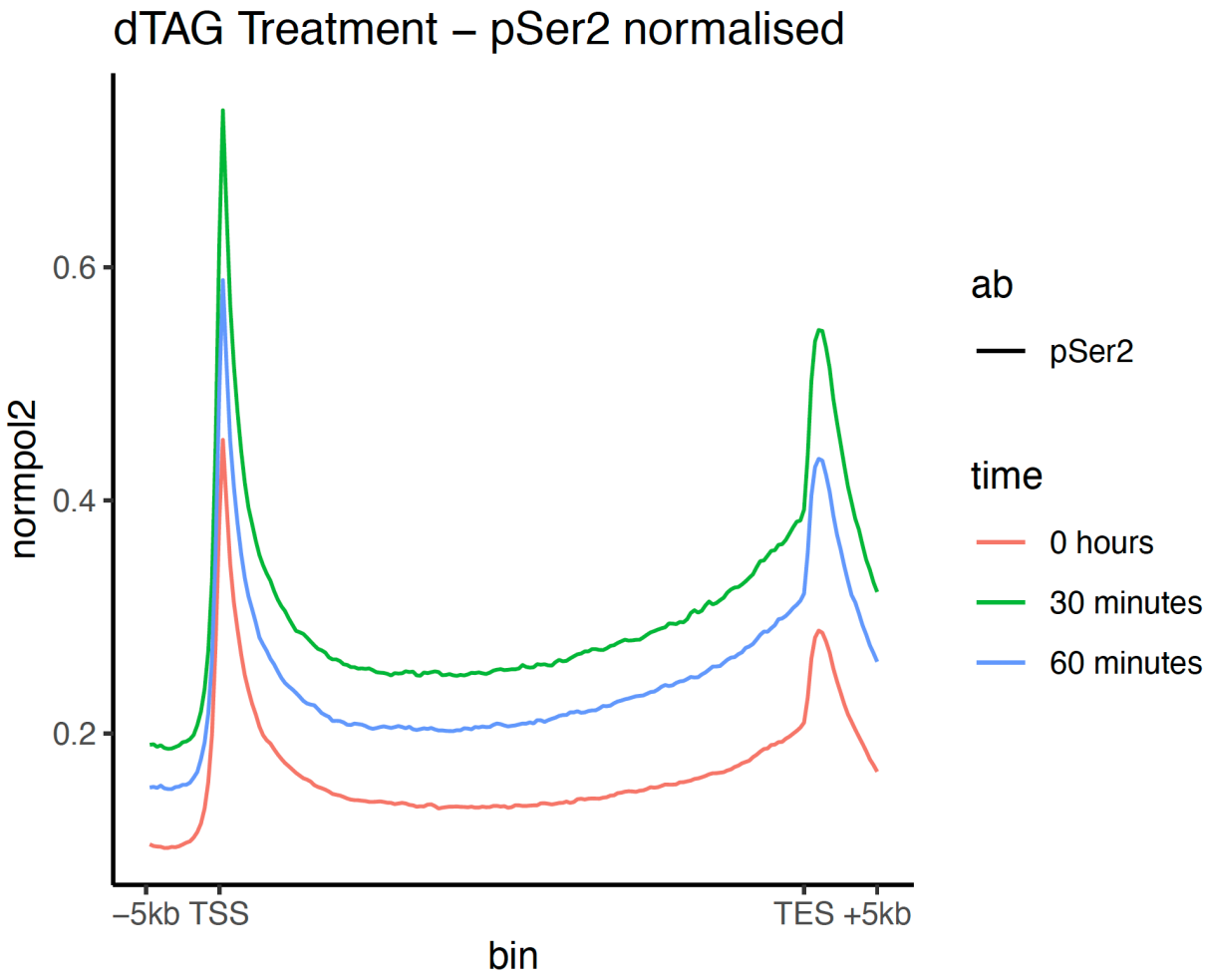



Figure 30: Average profile of pSer2 phosphorylation across the transcription start site (TSS), gene body, and transcription end site (TES) in clone 10 cells. The graph displays the normalized mean density of RNA Polymerase II (RNAPII) phosphorylated at serine 2 (pSer2) at three time points: 0 hours, 30 minutes, and 60 minutes post-dTAG treatment. The x-axis represents the distance from the TSS, covering -5 kb upstream of the TSS to +5 kb downstream of the TES. The y-axis shows the normalized pSer2 signal. While the baseline pSer2 signal is observed at 0 hours, there is a significant increase in pSer2 levels at 30 minutes and 60 minutes after treatment across the TSS, gene body, and TES regions. This increase suggests that, despite a reduction in overall RNAPII occupancy, the RNAPII that remains is more heavily phosphorylated at serine 2, indicating an enhanced transcriptional elongation activity in response to INTS3 ablation.

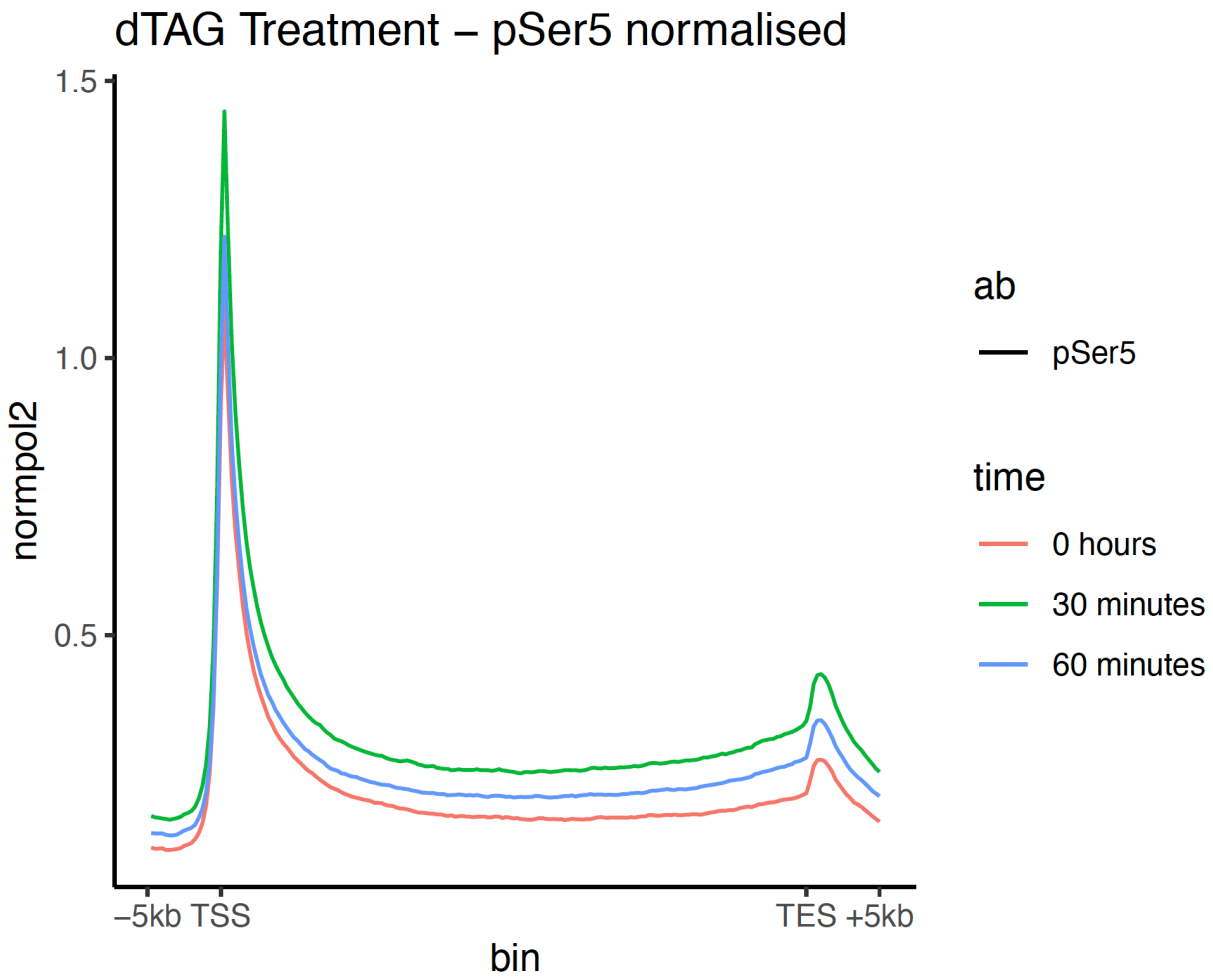



Figure 31: Average profile of pSer5 phosphorylation across the transcription start site (TSS), gene body, and transcription end site (TES) in clone 10 cells. The graph displays the normalized mean density of RNA Polymerase II (RNAPII) phosphorylated at serine 5 (pSer5) of the C-terminal domain at three time points: 0 hours, 30 minutes, and 60 minutes post-treatment. The x-axis represents the distance from the TSS, spanning from -5 kb upstream of the TSS to +5 kb downstream of the TES. The y-axis shows the normalized pSer5 signal. The pSer5 signal is strongly enriched at the TSS at all time points, indicating active transcription initiation. Notably, there is an increase in pSer5 levels at 30 minutes and 60 minutes compared to 0 hours, suggesting that the treatment enhances both transcription initiation and early elongation phases.

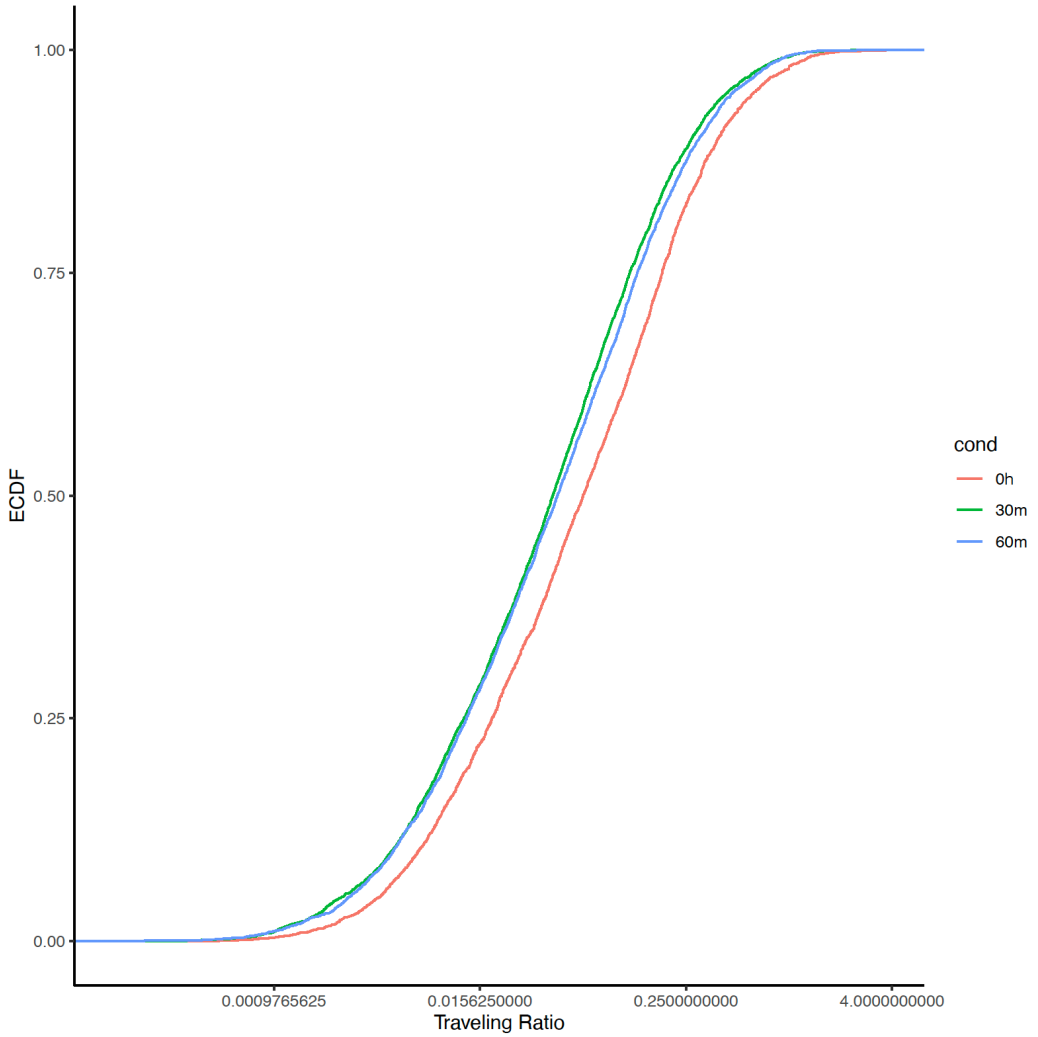



Figure 32: Traveling ratio of RNAPII after dTAG treatment of 0, 30 and 60 minutes in clone 10 cells. The ECDF curves represent RNAPII traveling ratios measured under 0h, 30m, and 60m treatment conditions. The leftward shift in the ECDF curves for the 30m and 60m treatments compared to the 0h treatment indicates a reduction in traveling ratios as treatment duration increases.

# *Chapter 3: Discussion*

INTS3 has been implicated in DNA damage recognition and repair alongside the SOSS complex, comprising additionally of NABP1/2 and INIP (Ren *et al.*, 2014). In addition, the transient docking of INTS3 to the Integrator complex prevents reassociation of RNA Pol II during premature transcriptional termination facilitated by Integrator (Fianu *et al.*, 2024). However, in order functionally address the role of INTS3 in key transcriptional processes on short time scales there was a need to establish a rapid degradation system for INTS3. A dTAG-based targeted protein degradation (TPD) system was developed in order to elicit precise temporal control over INTS3 levels, which, hence, facilitated an investigation into its functional role in transcriptional regulation over short time scales. In this light, it was explored whether INTS3 plays a role in transcriptional regulation, particularly in its interactions with RNA polymerase II (RNAPII) and other components of the transcriptional machinery such as Integrator and the SOSS complex. Results obtained suggest that INTS3 plays a crucial role in maintaining proper transcriptional regulation and RNAPII processivity, with broader implications involving transcriptional dynamics. Furthermore, these results reveal that INTS3 occupies a definitive functional role alongside the Integrator and SOSS complexes.

Significant changes in the expression of various Integrator subunits following rapid INTS3 ablation was observed by Western blot and IP-MS. This may suggest that INTS3 plays a role in contributing to the integrity of the Integrator complex. Following the degradation of INTS3 there was a notable upregulation of INTS6, INTS1 and INTS5. This may indicate that, in the absence of INTS3, these subunits play compensatory roles in maintaining Integrator integrity and function. However, it must be noted that the impact of INTS3 ablation may be limited to particular functional modules (specifically the phosphatase and scaffolding module) within the Integrator complex as evident by the absence of upregulation for INTS11 and INTS10, which are constituents of the endonucleolytic and enhancer modules, respectively.

The modular nature of the Integrator complex is exemplified by the differential expression of INTS6, INTS1 and INTS5, and lack of upregulation of INTS11 and INTS10 observed by Western blot. Indeed, subunits in close architectural proximity to a missing subunit may need to be

upregulated or stabilized following the loss of a closely localized subunit to maintain complex integrity. The aforementioned subunits, namely INTS6, INTS5 and INTS1 exist in close proximity to INTS3 in a structural context, comprising part of the phosphatase and scaffold modules, respectively, and their upregulation suggests that their respective modules may be particularly sensitive to the loss of INTS3. This may be as a result of their roles in maintaining the structural integrity and function of the complex. Changes in the macromolecular assembly of the Integrator complex as evidenced by glycerol gradient following INTS3 ablation is shown by a shift in the fractionation pattern of INTS1 and PPP2R1A, a subunit of the PP2A phosphatase complex. INTS3 may, hence, play a role in the recruitment or stabilization of these subunits within INTAC. The absence of PPP2R1A in the higher molecular weight fraction following INTS3 ablation may indicate that INTS3 may be involved in the assembly or stability of the phosphatase module of INTAC, further emphasizing its contribution to complex structural integrity and transcriptional regulation. Furthermore, in terms of complex composition, ChIP-seq data indicates that the SOSS complex (specifically INTS3 and NABP1) is present alongside RNAPII at the promoter regions of actively transcribed genes; hence the SOSS complex may be a component of INTAC.

The analysis of the differential expression changes for the transcriptome contributed additional evidence for the role of INTS3 in transcriptional regulation. Following INTS3 ablation, differential gene expression analysis revealed significant changes in the expression of immediate early response genes (IERGs). IERGs are genes that are precipitously induced in response to various stimuli, such as stress, growth factors and other signaling events. The upregulation of genes such as EGR1, FOS, JUND, and FOSB following short-term INTS3 degradation suggests that INTS3 may play a role in modulating the expression of these genes under normal conditions.

Additionally, the upregulation of IERGs may be a consequence of the increase of pSer2 and pSer5 levels for RNAPII across the genome resulting from rapid INTS3 ablation, as observed by ChIP-seq. Increased levels of pSer2 are specifically associated with RNAPII engagement in productive elongation. As IERGs are persistently poised for transcription with RNAPII existing in a paused conformation at their promoter regions (Saha *et al.*, 2011), an observed increase in pSer2 levels may contribute to an increase in their transcription. Furthermore, the fact that IERGs are upregulated in the absence of INTS3 may suggest that INTS3 functions as a transcriptional

repressor or modulator for these genes. Moreover, the enrichment of INTS3 and NABP1 at promoter and enhancer regions across the genome alongside RNAPII, as revealed by ChIP-seq analysis, as well as being identified to localize strongly to the AP-1 motif suggests that INTS3 may be involved in enhancer-promoter interactions that regulate the expression of IERGs.

INTS3 ablation significantly affected RNAPII dynamics, in terms of RNAPII occupancy and phosphorylation states, across the genome. In particular, following INTS3 degradation, it was observed that RNAPII density at both the transcription start site (TSS) and along the gene body was reduced. This suggests that INTS3 is essential for the proper retention and recruitment of RNAPII during transcription. Such a reduction in RNAPII occupancy may be indicative of INTS3 involvement in facilitating RNAPII transition from initiation to productive elongation – a critical step in ensuring efficient gene expression.

The increase of pSer2 and pSer5 phosphorylation states of RNAPII following INTS3 ablation provides additional insights into the functional role of INTS3. The phosphorylation of the carboxy-terminal domain (CTD) of RNAPII at Ser2 is typically associated with productive elongation, while phosphorylation at Ser5 is associated with transcriptional initiation. The observed increase in these phosphorylation states, despite a reduction in overall RNAPII occupancy, suggests that the remaining RNAPII complexes may enter a hyperphosphorylated state in response to INTS3 degradation. This hyperphosphorylation could imply that the absence of INTS3 deleteriously affects the phosphatase module of INTAC, with PP2A-C unable to dephosphorylate the CTD substantially, hence causing hyperphosphorylation and thus enhancing the elongation and initiation phases of transcription. This result coincides with a previous study that has highlighted the importance of CTD phosphorylation in regulating RNAPII processivity and transcriptional output. Ahn *et al*. (2004) demonstrated that the phosphorylation of the CTD is tightly regulated, and this regulation is fundamental to the transitioning of RNAPII from initiation to elongation. As the levels of pSer2 and pSer5 were observed to be increased at the promoter and along the gene body across the 5000 most active genes in OVCAR8, it may suggest that INTS3 plays a role in modulating CTD phosphorylation. This may further influence the ability of RNAPII to progress through the transcription cycle. Furthermore, this modulation of RNAPII dynamics may represent a novel mechanism in which INTS3, through the Integrator complex, acts to regulate gene expression.

Furthermore, the traveling ratio, which is calculated as the ratio of RNAPII present at the TSS over RNAPII present in the gene body, and is a measure used to assess transcriptional processivity, is significantly decreased in the absence of INTS3 over short time scales. This suggests that INTS3 plays a critical role in facilitating the processivity of RNAPII during transcription. The observed decrease in the traveling ratio after INTS3 ablation implies that the elongation phase of RNAPII is increasingly affected as INTS3 absence persists. Indeed, this may suggest that INTS3 has an essential function in the maintenance of efficient transcriptional elongation, whereby ablation of INTS3 gives rise to a predominance of RNAPII in the gene body, possibly as a result of disruptions to INTAC.

The result, following dTAG treatment and subsequent INTS3 ablation, that there is overall decreased RNAPII across the genome alongside a decreased traveling ratio coupled with an increase in pSer2 phosphorylation state suggests a complex transcriptional response to INTS3 depletion. One possible explanation is that INTS3 loss disrupts RNAPII recruitment at promoters, leading to overall reduced occupancy while already elongating polymerases continue with transcription. Alternatively, increased premature termination could reduce RNAPII density while some polymerases escape termination, thus increasing pSer2 levels for those polymerases. Furthermore, given the suspected interaction between INTS3 and PP2A, there may be a loss of PP2A-mediated dephosphorylation leading to sustained pSer2 phosphorylation despite a global decline in RNAPII occupancy.

Overall, these results implicate INTS3 in performing a broad array of functions including modulating RNAPII dynamics, regulating IERG expression, and contributing to the integrity of the Integrator complex as a whole. The use of a TPD system for INTS3 allowed for probing the rapid effects of INTS3 on transcriptional dynamics, differential transcriptomic expression and effect on overall Integrator complex stability. The upregulation of certain Integrator subunits following INTS3 ablation may imply that INTS3 plays a role in maintaining INTAC complex integrity, especially of the phosphatase and scaffold modules of INTAC. This is further indicated by the changes in the macromolecular assembly of the Integrator complex following INTS3 ablation, as evidence by different fractionation patterns following INTS3 degradation. Moreover,

INTS3 absence leads to the upregulation of various IERGs which may be a consequence of INTS3 modulating these genes under normal conditions, or instead, as a result of the increase in pSer2 and pSer5 of RNAPII leading to increased transcription. Furthermore, INTS3 degradation leads to a global increase in the levels of pSer2 and pSer5 for RNAPII, despite lower levels of total RNAPII. This suggests that, following INTS3 degradation, RNAPII becomes hyperphosphorylated either to compensate for lower overall RNAPII or as a result of INTS3 affecting the function of the phosphatase module which typically acts to dephosphorylate the CTD of RNAPII. Lastly, INTS3 ablation leads to a decrease in the traveling ratio of RNAPII. This implies that INTS3 plays a direct role in RNAPII processivity whereby RNAPII occupies the TSS to a lesser extent following INTS3 ablation.

## Rigor and Reproducibility

Experiments were conducted to rigorously and reproducibly address the research question of the role of INTS3 in transcriptional regulation. Several measures were implemented to ensure this was achieved. Firstly, negative controls were employed to ensure the observed results were a consequence of the experimental variable and not as a result of confounding factors. This includes: the use of loading controls (in the case of this work, GAPDH) for immunoblotting, to ensure that equivalent quantities of cell lysate was added per condition and to confirm that transfer efficiency was shared across the membrane; the use of IgG in co-immunoprecipitation, to verify the specificity of antibody-mediated protein interactions; and the use of wild-type OVCAR8 treated with $dTAG^V$-1 in differential RNA gene expression to confirm that the dTAG molecule was not eliciting a change in RNA expression levels, which should only be observed in cells engineered to degrade INTS3 via dTAG targeted protein degradation (in this work, clone 10 cells). Furthermore, immunoblots, co-immunoprecipitation, and glycerol gradients were repeated at least twice to ensure reproducibility.

Wild-type OVCAR8 and dTAG-positive clone 10 cell lines were routinely tested for mycoplasma contamination. Clone 10 cells were cultured and maintained in medium supplemented with a maintenance dose of antibiotics (puromycin, neomycin and blasticidin) to ensure that only a homogenous population expressing a stably integrated dTAG knock-in was present. In addition, upon thawing and culturing of preserved clone 10 cells, INTS3 degradation facilitated by dTAG treatment was tested via Western blot to ensure that the cell line functionally maintained the ability to ablate INTS3.

Data was analyzed using validated data processing methods implemented in R (version 4.2.2) and statistical significance was defined as $p < 0.05$. For RNA sequencing data, quality control was performed using FastQC, and low-quality reads were removed and read adapters trimmed with cutadapt, followed by alignment to the human genome (hg19). Volcano plots were used to visualize fold changes (log2(FC)) against statistical significance (-log10(FDR)). For ChIP-seq data, sequenced reads were aligned to the hg19 and the dm3 *D. melanogaster* reference genome using Burrows Wheeler Alignment (BWA) tool with the MEM algorithm. Samtools was used to remove PCR duplicates and MACS2 was used for peak calling. Average read density across defined

genomic intervals was computed using Deeptools, and data was visualized using ggrepel packages. In addition, ChIP-seq data for pSer2 and pSer5 was normalized to total RNAPII levels across the genome to account for variability in RNAPII occupancy across different genomic loci; without normalization, increases in pSer2 and pSer5 could simply reflect an increase in RNAPII as opposed to increased phosphorylation and, therefore, normalizing to total RNAPII helps distinguish changes in phosphorylation dynamics from changes in RNAPII recruitment.

The study includes some caveats: firstly, only OVCAR8 cells and the derived engineered dTAG clone 10 cells from OVCAR8 were used in this study, limiting the generalizability of the findings to other cell types. OVCAR8 cells were used for coherence with previous work completed by our research group, as certain other experiments probing the implications of INTS3 on transcriptional regulation were conducted using OVCAR8 as a model cell line. The OVCAR8 cell line was utilized by virtue of its relatively high accommodation for transfection, while being adaptable to gene editing without excessive cell lethality. Furthermore, OVCAR8 expresses all subunits of the Integrator, PP2A and SOSS complexes at appreciable levels permitting the study of the interplay of INTS3, Integrator and transcriptional dynamics. Moreover, OVCAR8 cells are suitable for use in experiments such as ChIP-seq, RNA-seq and proteomics – techniques that were critical to this study. An additional caveat is that only one clone (clone 10) was analyzed. This was a result of the difficulty in procuring a mutant cell line that degrades INTS3 rapidly and dependably. While engineering the clone, more than 400 clones were screened for degradation, all with varying levels, velocity and reliability of INTS3 ablation; it was only clone 10 that was found after extensive screening to nearly fully ablate INTS3 within two hours of dTAG treatment. It would hypothetically be possible to procure additional clones with the capacity for dTAG degradation of INTS3 but this would come at significant cost and time, and thus as an initial investigation into the role of INTS3 and transcriptional regulation, one clone was seen to suffice.

## Future Directions

A significant next step following this work is to implement strategies that address the caveats of the study, thereby strengthening the robustness and generalizability of the findings. Firstly, in order to overcome having utilized only OVCAR8 as a model cell line in this study, there should be an extension of use of different cell line models in order to more broadly assess the applicability of the findings. In this case, either alternative ovarian cancer derived cell lines could be used, such as OVCAR3, SK-OV-3, or IGROV-1, or more desirably, the study could be extended to include non-ovarian derived cell lines with similar molecular features such as high expression of Integrator, PP2A and SOSS subunits. Secondly, in order to address the limitation of having used only one clone (clone 10), several approaches could be considered: the most obvious and attractive, albeit less straightforward, approach would be to generate additional INTS3-dTAG positive clones with similar degradation kinetics to confirm reproducibility, first in the OVCAR8 cell line and following this in a different non-ovarian derived cell line. However, the difficulty and expense in procuring such a robust dTAG system must be restated, and thus alternative approaches may be more attractive to overcoming this caveat. Such strategies include complementary knock-out or knockdown approaches such as RNAi, CRISPR-based knockout, or an orthogonal approach such as Tet-on shRNA for INTS3. However, one would notably lose the rapid, targeted degradation elicited by the dTAG TPD system, and possibly the completeness of such degradation; in this light, results may become confounded or lead to various additional problems to be confronted such as off-target effects or incomplete ablation of INTS3. Therefore, while alternative approaches may be considered, the establishment of a dTAG system as an additional clone in OVCAR8 or a different cell line may be the most desirable strategy to address the caveat of only analyzing a single clone.

In addition to addressing the caveats of the study, further experiments may be conducted to comprehensively resolve the role of INTS3 in transcriptional regulation and thus make the findings appropriate for publication. Firstly, confirmation of the presence of INTS3 in the SOSS complex at sites of induced DNA DSBs should be undertaken to bolster the suggestion that the SOSS complex is crucial for the maintenance of genome integrity, and to observe the effect on RNAPII processivity. This could be accomplished by inducing DNA DSBs by virtue of ionizing radiation

or by an AsiSI-ER system which induces breaks in a site-specific manner, and then monitoring the resolution of the breaks and progression of RNAPII in wild-type OVCAR8 and clone 10 cells, with INTS3 ablated thus rendering the SOSS complex inactive, possibly by immunofluorescence monitoring gamma-H2AX, 53BP1 and RNAPII. By comparing wtOVCAR8 to clone 10 cells, this experiment would test whether INTS3 depletion disrupts the ability of RNAPII to traverse damage sites and continue transcription, thus implying an additional critical role of the SOSS complex in transcriptional regulation and strengthen the mechanistic link between INTS3, the SOSS complex, DNA repair and transcription. Furthermore, the effect of INTS3 ablation on R-loop (RNA-DNA hybrid structure) formation during transcription could be undertaken. In order to map R-loop formation, indicative of RNAPII accumulation and transcriptional hinderance, a technique such as MapR (Mutation and Pol II-associated chromatin Retention) could be considered. MapR functions by allowing for the isolation of genomic regions forming R-loops through exploiting catalytically inactive RNAse-H1 fused to micrococcal nuclease (MNase). If INTS3 plays a role in R-loop resolution or RNAPII processivity, MapR could reveal whether INTS3 depletion leads to an increase or decrease in R-loop accumulation at transcriptionally-active or DNA damage prone loci. Moreover, comparing wtOVCAR8 cells to clone 10 cells which express the INTS3 dTAG TPD system could determine the extent to which INTS3 is involved R-loop regulation or transcription restart after DNA damage.

The results of this work raise several questions for future research; one important area of investigation is the structural basis of INTS3 interactions within the Integrator complex and with other components of the transcriptional machinery. While it is known that INTS3 prevents the reassociation of RNAPII to the post-termination complex of Integrator (Fianu *et al.*, 2024), understanding the molecular details of the precise interactions of INTS3 with Integrator will be crucial for elucidating the definite mechanisms by which INTS3 influences transcriptional regulation. In addition, exploring the short-time course effects of INTS3 degradation on DNA damage accumulation should be undertaken, as the SOSS complex is implicated in DNA damage response.

Moreover, in order to further understand transcriptional dynamics and gene expression following INTS3 ablation, nascent transcript monitoring, possibly by transient transcriptome sequencing

(TT-seq), should be conducted. TT-seq would allow for the observation of transcriptional kinetics and gene regulatory mechanisms following INTS3 degradation which could provide additional insights to the contribution of INTS3 to transcriptional regulation.

Another important avenue for future research is the exploration of the role of INTS3 in different cell types and transcriptional contexts. While this research focused on the role of INTS3 in OVCAR8 cells, it is likely that INTS3 may have cell-type specific functions, such as an effect on cellular differentiation, particularly in tissues or cells where rapid transcriptional responses are critical, or in undifferentiated cell-types such as stem cells or induced pluripotent stem cells (iPSCs). Investigating the role of INTS3 in other cellular contexts could additionally provide a more comprehensive understanding of its functional repertoire, and its potential as a therapeutic target in diseases where transcriptional dysregulation is a key feature.

Furthermore, this work sets the stage for additional questions concerning INTS3 and its role in transcriptional regulation. One question pertinent to transcription is to understand how INTS3 depletion affects elongation rates of native transcripts and whether its absence alters the efficiency or processivity of RNAPII during transcriptional elongation. In order to resolve this question, native elongating transcript sequencing (NET-seq) could be conducted to assess RNAPII elongation kinetics following INTS3 degradation, as it provides single-nucleotide resolution of native RNA transcripts actively engaged with RNAPII. This technique is particularly well suited for detecting transcriptional dynamics in real-time, allowing for the identification of changes in RNAPII pausing, elongation rates and termination efficiency of RNAPII in the absence of INTS3. Additionally, NET-seq enables the mapping of transcriptional bottlenecks and sites of RNAPII stalling across the genome, thus offering crucial insights into whether INTS3 is required for maintaining proper RNAPII processivity and preventing aberrant transcriptional disruptions. Through the application of NET-seq in wild-type and INTS3-depleted conditions, it would be possible to determine whether INTS3 acts as a stabilizing factor for elongating RNAPII or plays a role in facilitating the transition between transcriptional states. Moreover, another question would be if whether INTS3 plays a role in the recruitment or regulation of transcription factors at immediate early-response gene (IERG) promoters which are rapidly and transiently activated in response to various stimuli and are important for regulating downstream transcriptional programs.

To address this question, ChIP-seq could be conducted for transcription factors such as AP-1 and EGR1 in INTS3-degraded versus control cells such as clone 10 cells untreated with the dTAG molecule to determine whether the absence of INTS3 affects transcription factor occupancy at these promoters. Additionally, knowing the extent to which INTS3 depletion influences enhancer-associated transcription would be important, as enhancers play a role in modulating gene expression by facilitating RNAPII recruitment and activation at target genes. This could be accomplished by conducting global run-on sequencing (GRO-seq) or a similar technique such as fastGRO which would enable the precise measurement of nascent transcription at enhancers and gene bodies, thus providing insights into whether INTS3 is required for maintaining proper enhancer transcriptional output. Furthermore, ChIP-seq for H3K27ac could be conducted to examine enhancer activity prior to and after INTS3 ablation to determine whether its depletion affects enhancer activation or interaction with transcriptional machinery. Finally, understanding the degree of functional redundancy of INTS3 with other Integrator subunits such as INTS6 and INTS8 should be explored. This redundancy may be understood through the generation of INTS6 and INTS8 knockout clones or possibly generating clones with a TPD system for INTS6 or INTS8; following this, the repetition of key experiments in this work with such clones, and the comparison of such results obtained with the INTS3 dTAG clone should be conducted.

# *Chapter 4: Methods*

**Reagents**

| **Reagent /Resource** | **Source** | **Identifier** |
|---|---|---|
| **Antibodies** | | |
| INT3 Antibody | Novus Biologicals | NBP1-19091 |
| DICE1 Antibody (H-6) | Santa Cruz | sc-376524 |
| Anti-INTS1 Antibody, clone 4.47 | Millipore Sigma | MABS1984 |
| INT11 Polyclonal Antibody | Bethyl Laboratories | A301-274A |
| INTS10 Polyclonal antibody | Proteintech | 15271-1-AP |
| Pol II Antibody (8WG16) | Santa Cruz | sc-56767 |
| RNA pol II CTD phospho Ser5 antibody (mAb) (3E8) | Active Motif | 61085 |
| INTS5 Polyclonal antibody | Proteintech | 14069-1-AP |
| RNA pol II CTD phospho Ser2 antibody (mAb) (3E10) | Active Motif | 61083 |
| RNA pol II CTD phospho Ser7 antibody (mAb) (3D4A12) | Active Motif | 61703 |
| PPP2CA Polyclonal antibody | Proteintech | 13482-1-AP |
| Anti-PP2A Antibody, C subunit, clone 1D6 | Millipore Sigma | 05-421 |
| OBFC2B Polyclonal antibody | Proteintech | 14809-1-AP |
| OBFC2A Polyclonal antibody | Proteintech | 16719-1-AP |
| HA Tag Monoclonal Antibody (2-2.2.14) | Thermo Fisher | 26183 |
| Spike-in Antibody (*D. melanogaster*) | Active Motif | 61686 |
| GAPDH (14C10) Rabbit mAb | Cell Signaling Technology | 2118L |
| PPP2R1A Polyclonal antibody | Proteintech | 15882-1-AP |
| Anti-mouse IgG, HRP-linked Antibody | Cell Signaling Technology | 7076S |
| Anti-rabbit IgG, HRP-linked Antibody | Cell Signaling Technology | 7074S |

| | | |
|---|---|---|
| Anti-rat IgG, HRP-linked Antibody | Cell Signaling Technology | 7077S |
| **Plasmids** | | |
| pSpCas9(BB)-2A-Puro (PX459) V2.0 | AddGene | 62988 |
| pCRIS-PITChv2-dTAG-Puro (BRD4) | AddGene | 91796 |
| pCRIS-PITChv2-dTAG-BSD (BRD4) | AddGene | 91795 |
| EasyFusion T2A-H2B-miRFP703 | AddGene | 113097 |
| INTS3-dTAG-Puro | This paper | NA |
| INTS3-dTAG-BSD | This paper | NA |
| INTS3-dTAG-Neo | This paper | NA |
| **Cell culture** | | |
| RPMI 1640, 1X | Corning | 10-040-CM |
| L-glutamine, 200 mM, 1X | Corning | 25-005-CI |
| Penicillin-Streptomycin (10,000 U/mL) | Thermo Fisher | 15140122 |
| 0.25% Trypsin, 2.21 mM EDTA, 1X | Corning | 25-053-CI |
| Fetal Bovine Serum (FBS) | Peak Serum | PS-FB3 |
| Puromycin | InvivoGen | ant-pr-1 |
| Blasticidin S HCl | Thermo Fisher | A1113903 |
| Neomycin (G418 sulfate) | Corning | 30-234-CR |
| dTAGV-1 hydrochloride | Tocris Biosciences | 7374 |
| Phosphate-Buffered Saline, 1X without calcium and magnesium, pH 7.4 ± 0.1 | Corning | 21-040-CM |
| **Biological and Chemical** | | |
| UltraPure Agarose | Thermo Fisher | 16500100 |
| 1 kb Plus DNA Ladder | NEB | N3200L |
| UltraPure™ TBE Buffer, 10X | Thermo Fisher | 15581028 |
| DNA Loading Dye & SDS Solution (6X) | Thermo Fisher | R1151 |
| Bio-Rad Protein Assay Dye Reagent Concentrate | Bio-rad | 5000006 |
| Bolt™ Bis-Tris Plus Mini Protein Gels, 4-12%, 1.0 mm, WedgeWell™ | Thermo Fisher | NW04120BOX |
| 10X Bolt™ Sample Reducing Agent | Thermo Fisher | B0009 |
| 4X Bolt™ LDS Sample Buffer | Thermo Fisher | B0007 |
| PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa | Thermo Fisher | 26620 |
| 20X Bolt™ MES SDS Running Buffer | Thermo Fisher | B000202 |
| 10x Tris/Glycine Buffer | Bio-rad | 1610734 |

| | | |
|---|---|---|
| Thick Blot Filter Paper, Precut, 7.5 x 10 cm | Bio-rad | 1703932 |
| Immun-Blot PVDF Membrane | Bio-rad | 1620177 |
| Bovine Serum Albumin | Sigma-Aldrich | A7906 |
| Clarity Western ECL Substrate | Bio-rad | 1705061 |
| Aprotinin from bovine lung | Sigma-Aldrich | A6279 |
| Leupeptin | Sigma-Aldrich | EI8 |
| Pepstatin | Sigma-Aldrich | 10253286001 |
| Dithiothreitol , Fisher BioReagents™ | Fisher Scientific | BP172-5 |
| Benzonase® Nuclease | Millipore Sigma | E1014 |
| Dynabeads™ Protein A for Immunoprecipitation | Thermo Fisher | 10002D |
| Dynabeads™ Protein G for Immunoprecipitation | Thermo Fisher | 10003D |
| Beckman Coulter AMPure XP | Fisher Scientific | NC9933872 |
| T4 Polynucleotide Kinase | New England Biosciences | M0201S |
| T4 Polynucleotide Kinase Reaction Buffer | New England Biosciences | B0201S |
| BbsI-HF® | New England Biosciences | R3539S |
| Quick CIP | New England Biosciences | M0525S |
| T4 DNA Ligase | New England Biosciences | M0202S |
| T4 DNA Ligase Reaction Buffer | New England Biosciences | B0202S |
| NEBuilder® HiFi DNA Assembly Master Mix | New England Biosciences | E2621S |
| Formaldehyde solution | Millipore Sigma | 252549-25ML |
| milliTUBE 1 ml AFA Fiber(100) | Covaris | 520130 |
| D5000 ScreenTape | Agilent | 5067-5588 |
| High Sensitivity D5000 DNA ScreenTape | Agilent | 5067-5592 |
| High Sensitivity D5000 Reagents | Agilent | 5067-5593 |
| High Sensitivity D5000 Ladder | Agilent | 5067-5594 |
| RNase A, DNase and protease-free (10 mg/mL) | Thermo Fisher | EN0531 |
| Proteinase K from *Tritirachium album* | Millipore Sigma | P2309-100MG |

| | | |
|---|---|---|
| Spike-in Chromatin (*D. melanogaster*) | Active Motif | 53083 |
| **Commercial Kits** | | |
| Neon™ Transfection System 100 µL Kit | Thermo Fisher | MPK10096 |
| Direct-zol RNA Miniprep Kit | Zymo Research | R2051 |
| Wizard® SV Gel and PCR Clean-Up System | Promega | A9281 |
| GeneJET Genomic DNA Purification Kit | Thermo Fisher | K0721 |
| GeneJET Plasmid Maxiprep Kit | Thermo Fisher | K0491 |
| ChIP DNA Clean & Concentrator | Zymo Research | D5205 |
| NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® | New England Biosciences | E7645S |

**Cell lines**

Wild-type OVCAR8 cells were cultured at 37 °C and 5 % $CO_2$ in Roswell Parks Memorial Institute (RPMI) 1640 medium (Corning) supplemented with 10 % fetal bovine serum (FBS) (Peak Serum), 2 mM L-glutamine (Corning), and 100 units/mL penicillin and 100 µg/mL streptomycin from a 100X stock (Corning). Clone 10 cells were cultured at 37 °C and 5 % $CO_2$ in RPMI 1640 medium supplemented with 10 % FBS, 2 mM L-glutamine,100 units/mL penicillin and 100 µg/mL streptomycin, 2 µg/mL puromycin (InvivoGen), 4 µg/mL blasticidin (Thermo Fisher Scientific), and 800 µg/mL G418 sulfate (neomycin) (Corning).

**Generation of CRISPR-Cas9 sgRNA and homology directed repair (HDR) plasmids**

*CRISPR-Cas9 sgRNA plasmids*

Oligonucleotides for annealing and ligation were ordered from Integrated DNA Technologies (IDT) for both sgRNA plasmids. For the first plasmid, the pair of forward oligo 5`-CAC CCC TTC AGA CTC TGA CCT GAG-3` and the reverse oligo 5`- AAA CCT CAG GTC AGA GTC TGA AGG-3` were used; for the second plasmid the pair of forward oligo 5`-CAC CTG GTG GAA AGG TCT CCA CTC-3` and reverse oligo 5`-AAA CGA GTG GAG ACC TTT CCA CCA-3`

were used. The spCas9 V2.0 plasmid vector (PX459) was ordered from AddGene (pSpCas9(BB)-2A-Puro (PX459) V2.0: #62988).

For the annealing of each oligo pair, 1μL of each oligo diluted to 100 μM, 1 μL of T4 polynucleotide kinase (PNK) (NEB), 1 μL of T4 PNK ligase reaction buffer (NEB) and 5 μL of dd$H_2O$ were mixed and incubated in a thermocycler at 37 °C for 30 minutes, then 95 °C for 5 minutes, followed by ramping down to 25 °C at an interval of 5 °C/min.

The PX459 vector was used to transform NEB 5α competent *E. coli* (NEB), then amplified and purified from bacterial exponential phase liquid culture using the GeneJET Plasmid Maxiprep Kit (Thermo Fisher Scientific). Three micrograms of the PX459 vector was linearized by restriction enzyme BbsI-HF (NEB) overnight at 37 °C followed by enzyme inactivation at 65 °C for 30 minutes. The linearized plasmid was dephosphorylated by adding 2 μL Quick CIP (calf intestinal alkaline phosphatase) (NEB) directly to the linearized plasmid solution and incubated at 37 °C for 30 minutes. The product was run on a 0.8 % agarose gel, with the plasmid band cut and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega) and analyzed for concentration by Nanodrop spectrophotometer (Thermo Scientific).

Ligation of the annealed oligos to the linearized plasmid was performed by first diluting the annealed sgRNA oligos 1 : 200 in dd$H_2O$, whereby 2 μL of the dilution was added to 1 μL of T4 DNA ligase (NEB), 2 μL T4 DNA ligase reaction buffer (NEB), 20 ng of linearized PX459 plasmid, and made up to 20 μL in dd$H_2O$ followed by incubation in a thermocycler at 25 °C for 15 minutes. The ligation reaction was used to transform NEB 5α competent *E. coli* (NEB), then colony selected for by ampicillin, followed by amplification and purification from bacterial exponential phase liquid culture using the GeneJET Plasmid Maxiprep Kit. Purified plasmids were validated to have the correct ligated insert by Sanger sequencing.

*Homology directed repair (HDR) plasmids*

Gibson assembly was performed for generation of HDR plasmids; four fragments were used in the assembly: a pBlueScript backbone, left homology arm (LHA) fragment, $FKBP^{F36V}$ insert, and a right homology arm (RHA) fragment. The pBlueScript backbone was amplified from an EasyFusion T2A-H2B-miRFP703 plasmid (AddGene: #113097), the LHA fragment was ordered as a gBlock (Integrated DNA Technologies), the $FKBP^{F36V}$ insert was amplified from a pCRIS-PITChv2-dTAG-Puro (BRD4) plasmid (AddGene: #91796) or a pCRIS-PITChv2-dTAG-BSD (BRD4) plasmid (AddGene: #91795), and the RHA fragment was amplified from gDNA extracted from wtOVCAR8 cells using a GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific). Each homology arm was designed to be 1 kb in length.

Polymerase chain reaction (PCR) was used to amplify the pBlueScript backbone, $FKBP^{F36V}$ insert, and RHA fragments prior to assembly. For amplification from plasmid DNA, 3 ng DNA was used, and for amplification from genomic DNA, 100 ng of DNA was used. The primers used for amplifying fragments prior to Gibson assembly were generated using NEBuilder online and ordered from Integrated DNA Technologies. Following PCR, fragments were gel purified on a 0.8 % agarose gel with the fragment band cut and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega) and analyzed for concentration by Nanodrop spectrophotometer (Thermo Scientific).

An equimolar ratio of fragments was used for Gibson assembly with 0.1 pmol of each fragment added to the reaction to ensure that the total mass of DNA did not exceed 0.5 pmol, to a total volume of 10 μL made up with $ddH_2O$. Fragments were added to 10 μL 2X NEBuilder® HiFi DNA Assembly Master Mix (NEB) and incubated at 50 °C for one hour. The Gibson assembly reaction was used to transform NEB 5α competent *E. coli* (NEB), then colony selected for by ampicillin, followed by amplification and purification from bacterial exponential phase liquid culture using the GeneJET Plasmid Maxiprep Kit. Purified plasmids were validated to have the correct ligated insert by Sanger sequencing.

**Transfection of OVCAR8 cells and colony picking**

Wild-type OVCAR8 cells, as well as subclones during subsequent introductions of HDR plasmids with differing antibiotic resistance cassettes, were transfected by electroporation using the Invitrogen™ Neon™ Transfection System (Invitrogen). Neon™ Transfection System 100 uL Kit (Invitrogen) pipette tips were loaded with cells at 20 x $10^6$ cells/mL, 8 μg HDR plasmid, and 12 μg of sgRNA plasmid, and electroporated at 1170 volts, 30 ms square waveform, and for two pulses. Following electroporation, cells were seeded in antibiotic-free RPMI 1640 medium supplemented with 10 % FBS and 2 mM L-glutamine and left to culture overnight. Medium was replaced the following day with fresh medium containing 100 units/mL penicillin and 100 μg/mL streptomycin.

Cells were left to grow for three days before addition of selection antibiotic, either 2 μg/mL puromycin (InvivoGen), 4 μg/mL blasticidin (Thermo Fisher), and/or 800 μg/mL G418 sulfate (Corning) in RPMI 1640 medium supplemented with 10 % FBS and 2 mM L-glutamine. Cells were grown for one week in antibiotic selection medium, with media refreshed every 2 – 3 days. Viable colonies were selected by scraping and replating via inverted microscope, and further expanded as individual clones. Clones were maintained in RPMI 1640 medium supplemented with 10 % FBS and 2 mM L-glutamine containing relevant antibiotics.

**SDS-PAGE and immunoblotting**

Cells were collected by trypsinization for 5 minutes, washed twice with ice cold phosphate buffered saline (PBS), and pelleted at 4 500 x *g*. Pellets were resuspended and lysed in cold lysis buffer (150 mM NaCl, 1% Triton X-100, 0.7% SDS, 500 mM DTT, 10 mM Tris-HCl, 5 mM EDTA) supplemented with 1 μg/mL aprotinin, 1 μg/mL leupeptin and 1 μg/mL pepstatin for 20 minutes. Lysed cells were centrifuged at 20 000 x *g* for 20 minutes at 4 °C and the supernatant harvested as protein extract. Protein concentration was determined by standard curve using Bio-Rad Protein Assay Dye Reagent Concentrate (Bio-rad). 4X Bolt™ LDS Sample Buffer (Thermo Fisher Scientific) and 10X Bolt™ Sample Reducing Agent (Thermo Fisher Scientific) were added

to 30 µg protein extracts, pipetted vigorously to mix, boiled at 95 °C for 5 minutes, and centrifuged briefly at 16 000 x *g*. Samples were loaded onto a Bolt™ Bis-Tris Plus Mini Protein Gel, 4 – 12 % gradient (Thermo Fisher Scientific) in Bolt™ MES SDS Running Buffer (Thermo Fisher Scientific), and separated by gel electrophoresis at 140 V until resolved. Loaded proteins were transferred to an Immun-Blot PVDF membrane at 90 V for 1,5 hours on ice, and following transfer, the membrane was incubated at room temperature for one hour in 10 %w/v bovine serum albumin (BSA) in Tris-buffered saline with Tween 20 (TBST) (Thermo Fisher Scientific). Incubation with primary antibody was conducted according to the manufacturer's instructions for antibody dilution factor in 5 %w/v BSA in TBST at room temperature for two hours, or overnight at 4 °C. Membranes were washed three times in TBST for 15 minutes, followed by incubation with secondary antibody conjugated to HRP at a dilution of 1 : 10000 in 5 %w/v BSA in TBST at room temperature. Membranes were washed three times in TBST for 5 minutes per wash and visualized using Clarity Western ECL (Bio-rad) on an Amersham Imager 680 (GE Healthcare).

**Co-immunoprecipitation**

Cells were collected by trypsinization for 5 minutes, washed twice with ice cold PBS, and pelleted at 4 500 x *g*. Pellets were resuspended and lysed in co-immunoprecipitation lysis buffer (1 mM PMSF, 10 mM NaF, 50 mM NaCl, 1 % Triton-X100 in PBS) supplemented with 1 µg/mL aprotinin, 1 µg/mL leupeptin and 1 µg/mL pepstatin, as well as 1 %v/v Benzonase nuclease (Sigma Aldrich). Volume of lysis buffer added was equivalent to 4X the volume of the pellet. Whole cell lysate was incubated for 30 minutes at 4 °C with rotation. Following incubation, lysates were centrifuged at 20 000 x *g* at 4 °C for 20 minutes. Supernatant was harvested and cleared with a needle and syringe. Protein concentration was determined by standard curve using Bio-Rad Protein Assay Dye Reagent Concentrate (Bio-rad). Depending on the antibody for immunoprecipitation, Dynabeads™ Protein A (Thermo Fisher Scientific) or Dynabeads™ Protein G (Thermo Fisher Scientific) were used; beads were prepared by washing twice with ice cold PBS followed by once in ice cold lysis buffer prior to use. Primary antibody was conjugated to beads in lysis buffer by incubation of antibody and beads at room temperature for 5 minutes with rotation, followed by incubation at 4 °C for 45 minutes with rotation. Following primary antibody conjugation, lysis

buffer was removed and 2 mg whole cell lysate was added to beads, topped up with lysis buffer. Co-immunoprecipitation was performed overnight at 4 °C with rotation. After co-immunoprecipitation, supernatant was discarded and beads washed three times with equal volumes of ice-cold co-immunoprecipitation wash buffer (150 mM NaCl, 0.05 % Tween-20 in dd$H_2O$). Proteins were eluted from beads in acidic glycine (0.1 M glycine, pH = 2.5) with agitation at 1 200 RPM for 2 minutes. Eluate was neutralized with Tris-HCl (pH = 9.5) and prepared for SDS-PAGE and immunoblotting according to the aforementioned method.

**Nuclear extraction**

Cells were collected by trypsinization for 5 minutes and trypsin inactivated with RPMI 1640 media supplemented with 10 % FBS and 2 mM L-glutamine. Cells were centrifuged at 3 000 RPM for 8 minutes at 4 °C. Supernatant was removed, and cells resuspended in ice cold PBS, followed by centrifugation at 3 000 RPM for 8 minutes at 4 °C. Once again, supernatant was removed and cells resuspended in ice cold PBS, followed by centrifugation at 3 000 RPM for 10 minutes at 4 °C. The packed cell volume (PCV) was noted and 5 PCVs of Buffer A (10 mM Tris-HCl at pH = 8, 1.5 mM $MgCl_2$, 10 mM KCl, 0.2 mM PMSF, 0.5 mM DTT) supplemented with 1 µg/mL aprotinin, 1 µg/mL leupeptin and 1 µg/mL pepstatin was used to gently resuspend the pellet. Cells were swelled for 10 minutes at 4 °C with rotation. Cells were then centrifuged at 2 000 RPM for 10 minutes at 4 °C. Supernatant was removed and 2 PCVs of Buffer A used to resuspend the pellet. Cells were homogenized by Dounce pestle in 12 strokes and homogenate then centrifuged at 2 500 RPM for 10 minutes at 4 °C. Supernatant was collected as the cytosolic fraction and the pellet resuspended in 1 PCV Buffer C (20 mM Tris-HCl at pH = 8, 1.5 mM $MgCl_2$, 0.42 M NaCl, 0.2 mM EDTA, 25 % glycerol, 0.2 mM PMSF, 0.5 mM DTT) supplemented with 1 µg/mL aprotinin, 1 µg/mL leupeptin and 1 µg/mL pepstatin. Resuspension was once again homogenized by Dounce pestle in 12 strokes followed by incubation for 30 minutes at 4 °C with rotation, followed by centrifugation for 30 minutes at 12 000 RPM at 4 °C. Supernatant was then dialyzed against nuclear extract dialysis buffer BC80 (20 mM Tris-HCl at pH = 8, 80 mM KCl, 0.2 mM EDTA at pH = 8, 10 % glycerol, 0.2 mM PMSF, 0.5 DTT) supplemented with 1 µg/mL aprotinin, 1 µg/mL leupeptin and 1 µg/mL pepstatin, for 18 hours at 4 °C. Dialysate was centrifuged for 30 minutes at 12 000 RPM

at 4 °C and the supernatant harvested as nuclear extract. Protein concentration was determined by standard curve using Bio-Rad Protein Assay Dye Reagent Concentrate (Bio-rad). Nuclear extracts were aliquoted, flash frozen in liquid nitrogen, and stored at -80 °C for downstream experiments.

**Glycerol gradient**

Glycerol gradients were performed using nuclear extracts as prepared according to the method for nuclear extraction. Nuclear extracts were cleared at 20 000 x *g* for 20 minutes at 4 °C. An 11 – 50 % glycerol gradient was prepared in 3 % glycerol increments in HEMG buffer (50 mM HEPES, 0.2 mM EDTA, 30 mM $MgCl_2$, and 200 mM KCl, 0.5 mM DTT) supplemented with 1 µg/mL aprotinin, 1 µg/mL leupeptin and 1 µg/mL pepstatin, and 1 mg nuclear extract added to the head of the gradient. Samples were centrifuged in a Thermo Scientific Sorvall WX+ Ultracentrifuge (Thermo Fisher Scientific) in a AH-650 rotor (Thermo Fisher Scientific) at 48 000 RPM for 16 hours at 4 °C. Fractions were collected and protein was precipitated by addition of 10 %v/v trichloroacetic acid (TCA) and incubation overnight at -20 °C. Fractions were centrifuged at 20 000 x *g* or 30 minutes at 4 °C. Pellets were washed twice in ice cold acetone and centrifuged at 20 000 x *g* for 10 minutes at 4 °C. Pellets were dried on ice for 3 hours and resuspended in 4X Bolt™ LDS Sample Buffer (Thermo Fisher Scientific) and 10X Bolt™ Sample Reducing Agent (Thermo Fisher Scientific) followed by vortexing. Proteins were boiled for 10 minutes at 95 °C and resolved on a Bolt™ Bis-Tris Plus Mini Protein Gel, 4 – 12 % gradient (Thermo Fisher Scientific) in Bolt™ MES SDS Running Buffer (Thermo Fisher Scientific), and separated by gel electrophoresis at 140 V until resolved. Immunoblotting was conducted according to the aforementioned method.

**IP followed by mass spectrometry**

Immunoprecipitation was performed using nuclear extracts prepared according to the aforementioned method. For each immunoprecipitation reaction, 1.5 mg of nuclear extract was used; 70 µL Dynabeads™ Protein A were prepared by washing in Co-IP buffer (20 mM Tris-HCl

at pH = 7.9, 100 mM NaCl, 0.1 % NP-40, 0.5 mM DTT) supplemented with 1 µg/mL aprotinin, 1 µg/mL leupeptin and 1 µg/mL pepstatin, followed by incubation with nuclear extract and 6 µg of primary antibody, overnight, at 4 °C with rotation. Following immunoprecipitation, beads were washed three times with 1 volume of Co-IP buffer and once with ice cold PBS + 0.05 % NP-40. Proteins were eluted from beads in acidic glycine (0.1 M glycine, pH = 2.5) with agitation at 1 200 RPM for 2 minutes. Eluate was neutralized with Tris-HCl (pH = 9.5). Eluates were prepared for SDS-PAGE in 4X Bolt™ LDS Sample Buffer (Thermo Fisher Scientific) and 10X Bolt™ Sample Reducing Agent (Thermo Fisher Scientific) and proteins resolved on a Novex™ Tris-Glycine Mini Protein Gel, 10%, in Tris/Glycine/SDS buffer (Bio-rad) at 100 volts for 15 minutes. Gels were stained overnight with Colloidal Blue Staining Kit (Invitrogen) and processed. The gel lanes were excised and digested by trypsin and analyzed by LC-MS/MS on a Q-Exactive Plus Mass Spectrometer. MS/MS spectra were searched with full tryptic specificity against the UniProt human database using the MaxQuant. The iBAQ values were normalized to the total levels of the protein that was pulled down.

**Quant-seq**

Cells were treated with dTAG$^{V}$-1 hydrochloride (Tocris) for 30 minutes, 1 hour and 4 hours before collection by trypsinization for 5 minutes and trypsin inactivated with RPMI 1640 media supplemented with 10 % FBS and 2 mM L-glutamine. Cells were lysed in an appropriate volume of TRI reagent (Zymo Research) and RNA was purified using the Direct-zol RNA Miniprep Kit (Zymo Research) with DNAse treatment. Barcoded RNA sequencing libraries were produced with the NEBNext Ultra Directional RNA Library Prep kit (NEB). RNA was sequenced on an Illumina NextSeq 2000 with 50 bp paired end reads. QC was performed on FASTQ files using FASTQC. Read adapters were trimmed using cutadapt and aligned to the hg19 human reference genome using STAR-2 v2.5 (quantMode TranscriptomeSAM --outFilterMultimapNmax 10 --outFilterMismatchNmax 10 --outFilterMismatchNoverLmax 0.3 --alignIntronMin 21 --alignIntronMax 0 --alignMatesGapMax 0 --alignSJoverhangMin 5 -- runThreadN 12 --twopassMode Basic --twopass1readsN 60000000 --sjdbOverhang 100). BAM files were sorted and filtered based on alignment quality (q = 10) with Samtools v0.1.19. Feature counts were used

to count mapped reads for genomic features such as genes and exons. DESeq2 used to identify differentially expressed genes using read counts normalized by feature length in RStudio 4.1.3. All plots were visualized with the ggplot package.

**Gene ontology analysis**

Gene Ontology (GO) analysis was conducted utilizing the Enrichr web-based tool (https://maayanlab.cloud/Enrichr/). A list of differentially expressed genes (DEGs) was initially derived from RNA-seq data through the application of DESeq2. This gene list was subsequently uploaded to Enrichr for GO term enrichment analysis. From the Enrichr database, significantly enriched GO terms across the categories of Biological Processes, Molecular Functions, and Cellular Components was facilitated. Submission of the DEG list to the Enrichr server and selecting the pertinent GO databases for enrichment analysis was conducted. The results provided included adjusted p-values to correct for multiple testing, thereby ensuring the reliability and statistical significance of the identified enriched terms. The top enriched GO terms were visualized using Enrichr's graphical output features.

**Chromatin immunoprecipitation followed by sequencing (ChIP-seq)**

Cells were treated with dTAG$^{V}$-1 hydrochloride (Tocris) for 30 minutes or 1 hour, or left untreated, and collected by trypsinization for 5 minutes and trypsin inactivated with RPMI 1640 media supplemented with 10 % FBS and 2 mM L-glutamine. Cells were pelleted at 300 x *g* and resuspended in 10 mL RPMI 1640 media supplemented with 10 % FBS and 2 mM L-glutamine. Cells were then crosslinked at room temperature for 5 minutes in 1 % formaldehyde with gentle rotation. Crosslinking was quenched with 0.125 M glycine and incubated for 5 minutes at room temperature with gentle rotation. Cells were centrifuged at 1 300 RPM for 5 minutes at 4 °C and washed twice in ice cold PBS, flash frozen on dry ice and stored overnight at -80 °C. The frozen pellet was resuspended in 900 µL ChIP buffer (150 mM NaCl, 1% Triton X-100, 5 mM EDTA, 10 mM Tris-HCl at pH = 7.5, 0.5 mM DTT) supplemented with 1 µg/mL aprotinin, 1 µg/mL leupeptin and 1 µg/mL pepstatin and 0.7 – 0.8 % SDS. Lysates were then sonicated by

ultrasonication (Covaris Focused-ultrasonicator ME220) at 20 % duty cycle, 1000 cycles/burst, and 10 intensity for 8.5 – 10 minutes at 6 °C. Following sonication, in order to assess DNA fragmentation, 20 µL of each sonicated sample was brought to 100 µL with Tris/EDTA buffer with 1 %v/v SDS and 5 µL of RNAse and 20 mg/mL proteinase K was added. Samples were digested for 1 hour at 65 °C. Samples were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega) and chromatin fragmentation was assessed by TapeStation (Agilent Technologies TapeStation 4200). For the immunoprecipitation reaction, sheared chromatin was incubated with Dynabeads™ Protein A or Dynabeads™ Protein G, 10 – 20 µg of antibody, 2 µg *D. melanogaster* Spike-in antibody (Active Motif), and *D. melanogaster* Spike-in chromatin from 2 x $10^6$ formaldehyde-crosslinked *D. melanogaster* S2 cells (Active Motif). Incubation of the immunoprecipitation reaction was conducted overnight at 4 °C with gentle rotation. Following immunoprecipitation, beads were washed twice with mixed micelle wash buffer (150 mM NaCl, 20 mM Tris-HCl at pH = 8, 5 mM EDTA, 65 %w/v sucrose, 20 % Triton X-100, 20 % SDS), then twice with Buffer 500 (0.1% deoxycholate, 500 mM NaCl, 1M HEPES at pH = 7.5, 1 mM EDTA, 1 % Triton X- 100), and then twice with LiCl/Detergent buffer (0.5 % deoxycholate, 250 mM LiCl, 1mM EDTA, 0.5% IGEPAL CA-630, 10 mM Tris-HCl at pH = 8.0). Beads were washed once in TE buffer (10 mM Tris-HCl at pH = 7.5, 1 mM EDTA). Beads were resuspended in 240 µL Tris/EDTA buffer with 1 %v/v SDS and incubated at 65 °C for 10 minutes with agitation at 1 200 RPM to eluate chromatin-bound protein from beads. Elution was repeated for a total of two times and eluates were reverse crosslinked overnight at 65 °C. Protein was degraded by addition of 300 µg/mL proteinase K to samples and incubated at 65 °C for 1 hour. Immunoprecipitated DNA was purified using the ChIP DNA Clean & Concentrator Kit (Zymo Research). DNA concentration was quantified by Qubit™ fluorometric quantification using the Qubit™ dsDNA Quantification Assay Kit (Thermo Fisher Scientific) and the Qubit 4 Fluorimeter (Thermo Fisher Scientific). Sequencing libraries were prepared using NEBNext Ultra II DNA Library Prep Kit and 200 – 500 base-pair size selection was performed using a Pippin Prep System and sequenced on an Illumina NextSeq 500 or NextSeq 2000. Sequenced reads were aligned to the hg19 human or dm3 *D. melanogaster* reference genome using Burrows Wheeler Alignment tool (BWA) with the MEM algorithm, and the generated files were converted to BAM files using Samtools which was additionally used to remove PCR duplicates (rndup) from BAM files, and converted to BigWig files. MACS2 was used

for peak calling. Average read density across defined genomic intervals was computed using the Deeptools computeMatrix function. Data was visualized in RStudio with ggplot2 (v4.1.2) and ggrepel packages. The traveling ratio was calculated considering the coverage of RNA Polymerase II at the TSS and comparing it to the coverage at the gene body through TES.

# *References*


Adelman, K. and Lis, J.T., 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics*, *13*(10), pp.720-731

Albrecht, T.R. and Wagner, E.J., 2012. snRNA 3′ end formation requires heterodimeric association of integrator subunits. *Molecular and cellular biology*, 32(6), pp. 1112-1123

Albrecht, T.R., Shevtsov, S.P., Wu, Y., Mascibroda, L.G., Peart, N.J., Huang, K.L., Sawyer, I.A., Tong, L., Dundr, M. and Wagner, E.J., 2018. Integrator subunit 4 is a 'Symplekin-like'scaffold that associates with INTS9/11 to form the Integrator cleavage module. *Nucleic acids research*, *46*(8), pp.4241-4255.

Baillat, D., Hakimi, M.A., Näär, A.M., Shilatifard, A., Cooch, N. and Shiekhattar, R., 2005. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell*, *123*(2), pp.265-276.

Barbieri, E., Trizzino, M., Welsh, S.A., Owens, T.A., Calabretta, B., Carroll, M., Sarma, K. and Gardini, A., 2018. Targeted enhancer activation by a subunit of the integrator complex. *Molecular cell*, *71*(1), pp.103-116.

Bartek, J. and Lukas, J., 2007. DNA damage checkpoints: from initiation to recovery or adaptation. *Current opinion in cell biology*, *19*(2), pp.238-245.

Bensimon, A., Pizzagalli, M.D., Kartnig, F., Dvorak, V., Essletzbichler, P., Winter, G.E. and Superti-Furga, G., 2020. Targeted degradation of SLC transporters reveals amenability of multi-pass transmembrane proteins to ligand-induced proteolysis. *Cell chemical biology*, *27*(6), pp.728-739

Buis, J., Wu, Y., Deng, Y., Leddon, J., Westfield, G., Eckersdorff, M., Sekiguchi, J.M., Chang, S. and Ferguson, D.O., 2008. Mre11 nuclease activity has essential roles in DNA repair and genomic stability distinct from ATM activation. *Cell*, *135*(1), pp.85-96.

Camlin, N.J. and Evans, J.P., 2019. Auxin-inducible protein degradation as a novel approach for protein depletion and reverse genetic discoveries in mammalian oocytes. *Biology of reproduction*, *101*(4), pp.704-718.

Chen, J., Ezzeddine, N., Waltenspiel, B., Albrecht, T.R., Warren, W.D., Marzluff, W.F. and Wagner, E.J., 2012. 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*(12), pp.2148-2156.

Corden, J.L., 2013. RNA polymerase II C-terminal domain: tethering transcription to transcript and template. *Chemical reviews*, *113*(11), pp.8423-8455.

Cossa, G., Parua, P.K., Eilers, M. and Fisher, R.P., 2021. Protein phosphatases in the RNAPII transcription cycle: erasers, sculptors, gatekeepers, and potential drug targets. *Genes & Development*, *35*(9-10), pp.658-676.

Dominski, Z., Yang, X.C., Purdy, M., Wagner, E.J. and Marzluff, W.F., 2005. A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Molecular and cellular biology*, *25*(4), pp.1489-1500.

Erb, M.A., Scott, T.G., Li, B.E., Xie, H., Paulk, J., Seo, H.S., Souza, A., Roberts, J.M., Dastjerdi, S., Buckley, D.L. and Sanjana, N.E., 2017. Transcription control by the ENL YEATS domain in acute leukaemia. *Nature*, *543*(7644), pp.270-274

Fianu, I., Chen, Y., Dienemann, C., Dybkov, O., Linden, A., Urlaub, H. and Cramer, P., 2021. Structural basis of Integrator-mediated transcription regulation. *Science*, *374*(6569), pp.883-887.

Fianu, I., Ochmann, M., Walshe, J.L., Dybkov, O., Cruz, J.N., Urlaub, H. and Cramer, P., 2024. Structural basis of Integrator-dependent RNA polymerase II termination. *Nature*, *629*(8010), pp.219-227.

Flanagan, P.M., Kelleher III, R.J., Sayre, M.H., Tschochner, H. and Kornberg, R.D., 1991. A mediator required for activation of RNA polymerase II transcription in vitro. *Nature*, *350*(6317), pp.436-438.

Fuchs, G., Voichek, Y., Benjamin, S., Gilad, S., Amit, I. and Oren, M., 2014. 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome biology*, *15*, pp.1-11.

Gromak, N., West, S. and Proudfoot, N.J., 2006. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Molecular and cellular biology*, *26*(10), pp.3986-3996.

Heidemann, M., Hintermair, C., Voß, K. and Eick, D., 2013. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1829*(1), pp.55-62.

Holland, A.J., Fachinetti, D., Han, J.S. and Cleveland, D.W., 2012. Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. *Proceedings of the National Academy of Sciences*, *109*(49), pp.E3350-E3357.

Hopkins, B.B. and Paull, T.T., 2008. The P. furiosus mre11/rad50 complex promotes 5′ strand resection at a DNA double-strand break. *Cell*, *135*(2), pp.250-260.

Hu, S., Peng, L., Xu, C., Wang, Z., Song, A. and Chen, F.X., 2021. SPT5 stabilizes RNA polymerase II, orchestrates transcription cycles, and maintains the enhancer landscape. *Molecular Cell*, *81*(21), pp.4425-4439

Huang, H.T., Seo, H.S., Zhang, T., Wang, Y., Jiang, B., Li, Q., Buckley, D.L., Nabet, B., Roberts, J.M., Paulk, J. and Dastjerdi, S., 2017. MELK is not necessary for the proliferation of basal-like breast cancer cells. *Elife*, *6*, p.e26693.

Huang, J., Gong, Z., Ghosal, G. and Chen, J., 2009. SOSS complexes participate in the maintenance of genomic stability. *Molecular cell*, *35*(3), pp.384-393.

Huang, K.L., Jee, D., Stein, C.B., Elrod, N.D., Henriques, T., Mascibroda, L.G., Baillat, D., Russell, W.K., Adelman, K. and Wagner, E.J., 2020. Integrator recruits protein phosphatase 2A to prevent pause release and facilitate transcription termination. *Molecular cell*, *80*(2), pp.345-358.

Inagaki, Y., Yasui, K., Endo, M., Nakajima, T., Zen, K., Tsuji, K., Minami, M., Tanaka, S., Taniwaki, M., Itoh, Y. and Arii, S., 2008. CREB3L4, INTS3, and SNAPAP are targets for the 1q21 amplicon frequently detected in hepatocellular carcinoma. *Cancer genetics and cytogenetics*, *180*(1), pp.30-36.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G. and Linsley, P.S., 2003. Expression profiling reveals off-target gene regulation by RNAi. *Nature biotechnology*, *21*(6), pp.635-637.

Jeronimo, C., Collin, P. and Robert, F., 2016. The RNA polymerase II CTD: the increasing complexity of a low-complexity protein domain. *Journal of molecular biology*, *428*(12), pp.2607-2622.

Jia, Y., Cheng, Z., Bharath, S.R., Sun, Q., Su, N., Huang, J. and Song, H., 2021. Crystal structure of the INTS3/INTS6 complex reveals the functional importance of INTS3 dimerization in DSB repair. *Cell Discovery*, *7*(1), p.66.

Jonkers, I. and Lis, J.T., 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nature reviews Molecular cell biology*, *16*(3), pp.167-177.

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

Kennedy, R.D. and D'Andrea, A.D., 2006. DNA repair pathways in clinical practice: lessons from pediatric cancer susceptibility syndromes. *Journal of Clinical Oncology*, *24*(23), pp.3799-3808

Kwak, H. and Lis, J.T., 2013. Control of transcriptional elongation. *Annual review of genetics*, *47*, pp.483-508.

Kwak, H., Fuda, N.J., Core, L.J. and Lis, J.T., 2013. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science*, *339*(6122), pp.950-953.

Lambrecht, C., Haesen, D., Sents, W., Ivanova, E. and Janssens, V., 2013. Structure, regulation, and pharmacological modulation of PP2A phosphatases. *Phosphatase Modulators*, pp.283-305

Li, J., Ma, X., Banerjee, S., Baruah, S., Schnicker, N.J., Roh, E., Ma, W., Liu, K., Bode, A.M. and Dong, Z., 2021. Structural basis for multifunctional roles of human Ints3 C-terminal domain. *Journal of Biological Chemistry*, *296*.

Li, M., Ball, C.B., Collins, G., Hu, Q., Luse, D.S., Price, D.H. and Meier, J.L., 2020. Human cytomegalovirus IE2 drives transcription initiation from a select subset of late infection viral promoters by host RNA polymerase II. *PLoS pathogens*, *16*(4), p.e1008402.

Liu, Y., Li, S., Chen, Y., Kimberlin, A.N., Cahoon, E.B. and Yu, B., 2016. snRNA 3′ end processing by a CPSF73-containing complex essential for development in Arabidopsis. *PLoS biology*, *14*(10), p.e1002571

Lok, T.M., Wang, Y., Xu, W.K., Xie, S., Ma, H.T. and Poon, R.Y., 2020. Mitotic slippage is determined by p31comet and the weakening of the spindle-assembly checkpoint. *Oncogene*, *39*(13), pp.2819-2834.

Malovannaya, A., Lanz, R.B., Jung, S.Y., Bulynko, Y., Le, N.T., Chan, D.W., Ding, C., Shi, Y., Yucer, N., Krenciute, G. and Kim, B.J., 2011. Analysis of the human endogenous coregulator complexome. *Cell*, *145*(5), pp.787-799

Malovannaya, A., Li, Y., Bulynko, Y., Jung, S.Y., Wang, Y., Lanz, R.B., O'Malley, B.W. and Qin, J., 2010. Streamlined analysis schema for high-throughput identification of endogenous protein complexes. *Proceedings of the National Academy of Sciences*, *107*(6), pp.2431-2436

Morawska, M. and Ulrich, H.D., 2013. An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast*, *30*(9), pp.341-351.

Moreno, R.Y., Juetten, K.J., Panina, S.B., Butalewicz, J.P., Floyd, B.M., Ramani, M.K.V., Marcotte, E.M., Brodbelt, J.S. and Zhang, Y.J., 2023. Distinctive interactomes of RNA polymerase II phosphorylation during different stages of transcription. *Iscience*, *26*(9).

Muhar, M., Ebert, A., Neumann, T., Umkehrer, C., Jude, J., Wieshofer, C., Rescheneder, P., Lipp, J.J., Herzog, V.A., Reichholf, B. and Cisneros, D.A., 2018. SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. *Science*, *360*(6390), pp.800-805.

Nabet, B., Ferguson, F.M., Seong, B.K.A., Kuljanin, M., Leggett, A.L., Mohardt, M.L., Robichaud, A., Conway, A.S., Buckley, D.L., Mancias, J.D. and Bradner, J.E., 2020. Rapid and direct control of target protein levels with VHL-recruiting dTAG molecules. *Nature communications*, *11*(1), p.4687.

Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A.L., Erb, M.A., Lawlor, M.A., Souza, A. and Scott, T.G., 2018. The dTAG system for immediate and target-specific protein degradation. *Nature chemical biology*, *14*(5), pp.431-441.

Natsume, T., Kiyomitsu, T., Saga, Y. and Kanemaki, M.T., 2016. Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. *Cell reports*, *15*(1), pp.210-218.

Ng, L.Y., Ma, H.T., Liu, J.C., Huang, X., Lee, N. and Poon, R.Y., 2019. Conditional gene inactivation by combining tetracycline-mediated transcriptional repression and auxin-inducible degron-mediated degradation. *Cell Cycle*, *18*(2), pp.238-248.

Nishimura, K. and Fukagawa, T., 2017. An efficient method to generate conditional knockout cell lines for essential genes by combination of auxin-inducible degron tag and CRISPR/Cas9. *Chromosome Research*, *25*, pp.253-260.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. and Kanemaki, M., 2009. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature methods*, *6*(12), pp.917-922.

Pfleiderer, M.M. and Galej, W.P., 2021. Structure of the catalytic core of the Integrator complex. *Molecular cell*, *81*(6), pp.1246-1259.

Prozzillo, Y., Delle Monache, F., Ferreri, D., Cuticone, S., Dimitri, P. and Messina, G., 2019. The true story of Yeti, the "Abominable" heterochromatic gene of Drosophila melanogaster. *Frontiers in physiology*, *10*, p.1093.

Prozzillo, Y., Fattorini, G., Santopietro, M.V., Suglia, L., Ruggiero, A., Ferreri, D. and Messina, G., 2020. Targeted protein degradation tools: overview and future perspectives. Biology (Basel) 9 (12).

Ren, W., Chen, H., Sun, Q., Tang, X., Lim, S.C., Huang, J. and Song, H., 2014. Structural basis of SOSS1 complex assembly and recognition of ssDNA. *Cell reports*, *6*(6), pp.982-991.

Richard, D.J., Bolderson, E., Cubeddu, L., Wadsworth, R.I., Savage, K., Sharma, G.G., Nicolette, M.L., Tsvetanov, S., McIlwraith, M.J., Pandita, R.K. and Takeda, S., 2008. Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. *Nature*, *453*(7195), pp.677-681.

Richter, W.F., Nayak, S., Iwasa, J. and Taatjes, D.J., 2022. The Mediator complex as a master regulator of transcription by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, *23*(11), pp.732-749.

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M. and Stainier, D.Y., 2015. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature*, *524*(7564), pp.230-233

Röth, S., Fulcher, L.J. and Sapkota, G.P., 2019. Advances in targeted degradation of endogenous proteins. *Cellular and molecular life sciences*, *76*(14), pp.2761-2777.

Sabath, K., Stäubli, M.L., Marti, S., Leitner, A., Moes, M. and Jonas, S., 2020. INTS10–INTS13–INTS14 form a functional module of Integrator that binds nucleic acids and the cleavage module. *Nature communications*, *11*(1), p.3422.

Saha, R.N., Wissink, E.M., Bailey, E.R., Zhao, M., Fargo, D.C., Hwang, J.Y., Daigle, K.R., Fenn, J.D., Adelman, K. and Dudek, S.M., 2011. Rapid activity-induced transcription of Arc and other IEGs relies on poised RNA polymerase II. *Nature neuroscience*, *14*(7), pp.848-856.

Schier, A.C. and Taatjes, D.J., 2020. Structure and mechanism of the RNA polymerase II transcription machinery. *Genes & development*, *34*(7-8), pp.465-488.

Seshacharyulu, P., Pandey, P., Datta, K. and Batra, S.K., 2013. Phosphatase: PP2A structural importance, regulation and its aberrant expression in cancer. *Cancer letters*, *335*(1), pp.9-18.

Sonoda, E., Hochegger, H., Saberi, A., Taniguchi, Y. and Takeda, S., 2006. Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. *DNA repair*, *5*(9-10), pp.1021-1029.

Taylor, E., Alqadri, N., Dodgson, L., Mason, D., Lyulcheva, E., Messina, G. and Bennett, D., 2017. MRL proteins cooperate with activated Ras in glia to drive distinct oncogenic outcomes. *Oncogene*, *36*(30), pp.4311-4322.

Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E. and Ljungman, M., 2014. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome research*, *24*(6), pp.896-905.

Vervoort, S.J., Welsh, S.A., Devlin, J.R., Barbieri, E., Knight, D.A., Offley, S., Bjelosevic, S., Costacurta, M., Todorovski, I., Kearney, C.J. and Sandow, J.J., 2021. The PP2A-Integrator-CDK9 axis fine-tunes transcription and can be targeted therapeutically in cancer. *Cell*, *184*(12), pp.3143-3162

Wang, Y., Wang, M., Zheng, T., Hou, Y., Zhang, P., Tang, T., Wei, J. and Du, Q., 2020. Specificity profiling of CRISPR system reveals greatly enhanced off-target gene editing. *Scientific reports*, *10*(1), p.2269.

Wang, Z., Zhang, C., Guo, J., Yang, Y., Li, P., Wang, Z., Liu, S., Zhang, L., Zeng, X., Zhai, J. and Wang, X., 2024. CRISPR-Cas9 screening identifies INTS3 as an anti-apoptotic RNA-binding protein and therapeutic target for colorectal cancer. *Iscience*, *27*(5).

Welsh, S.A. and Gardini, A., 2023. Genomic regulation of transcription and RNA processing by the multitasking Integrator complex. *Nature Reviews Molecular Cell Biology*, *24*(3), pp.204-220.

Young, B.A., Gruber, T.M. and Gross, C.A., 2002. Views of transcription initiation. *Cell*, *109*(4), pp.417-420.

Zhang, F., Ma, T. and Yu, X., 2013. A core hSSB1–INTS complex participates in the DNA damage response. *Journal of cell science*, *126*(21), pp.4850-4855.

Zhang, J. and Corden, J.L., 1991. Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *Journal of Biological Chemistry*, *266*(4), pp.2290-2296

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E. and Ebright, R.H., 2012. Structural basis of transcription initiation. *Science*, *338*(6110), pp.1076-1080.

Zheng, H., Qi, Y., Hu, S., Cao, X., Xu, C., Yin, Z., Chen, X., Li, Y., Liu, W., Li, J. and Wang, J., 2020. Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. *Science*, *370*(6520), p.eabb5872.

Zheng, H., Qi, Y., Hu, S., Cao, X., Xu, C., Yin, Z., Chen, X., Li, Y., Liu, W., Li, J. and Wang, J., 2020. Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. *Science*, *370*(6520), p.eabb5872

Zhou, Q., Li, T. and Price, D.H., 2012. RNA polymerase II elongation control. *Annual review of biochemistry*, *81*, pp.119-143.