

Abstract

Insights into the Biogenesis of Stress-Induced Readthrough Transcripts

Nicolle A. Rosa Mercado

2021

Readthrough transcription, defined as transcription extending beyond the annotated termination site of genes, is induced by cellular stress and results in the production of thousands of long noncoding RNAs called DoGs (downstream-of-gene transcripts). The localization and induction patterns of these transcripts have been extensively characterized in different cellular contexts. However, the mechanisms that lead to DoG production upon cellular stress remain unknown. To gain a better understanding of the mechanisms that lead to DoG production, I investigated the transcriptional and proteomic landscapes that accompany the induction of these RNAs in human cells exposed to hyperosmotic stress.

In collaboration with Dr. Matthew Simon's lab, I studied the transcriptional profiles that accompany DoG induction in the context of hyperosmotic stress using transient transcriptome (TT) sequencing combined with TimeLapse (TL) chemistry. The results revealed widespread transcriptional repression induced by hyperosmotic stress from which very few genes escape. Analyses of expressed genes that do not overlap with DoG regions demonstrate that DoGs are produced regardless of the transcriptional regulation of their upstream genes.

To assess how hyperosmotic stress impacts the protein interactome of RNA Polymerase (Pol) II, I collaborated with Dr. Jesse Rinehart's lab to perform mass

spectrometry and western blot analyses on anti-Pol II immunoprecipitates obtained from chromatin fractions of untreated and KCl-treated HEK-293T cells. Consistent with the transcriptional changes observed, mass spectrometry results revealed drastic alterations to the Pol II interactome upon stress. Interestingly, subunits of the Integrator complex were not detected among the Pol II interactors in cells exposed to hyperosmotic stress, while subunits of the cleavage and polyadenylation machinery were present in both untreated and stressed samples. ChIP-seq experiments performed with antibodies against Int11 and Int3 demonstrated a decreased occupancy of these subunits near the transcription start sites of DoG-producing genes and of non-DoG genes that was not dependent on a loss of Pol II binding at these sites. Furthermore, siRNA-mediated knockdowns of Int11, the catalytic subunit of the Integrator complex, revealed the induction of hundreds of DoGs. I compared the identity of genes that generate readthrough transcripts after Int11 knockdown to genes that produce DoGs after KCl-treatment and found an overlap of up to 25%. This suggests that the observed decrease of Int11 binding is important, but not solely responsible for DoG induction in the context of hyperosmotic stress.

The findings presented in this dissertation provide insights regarding the biogenesis of a recently characterized class of long noncoding RNAs in human cells exposed to hyperosmotic stress. My observations also clarify the extent of transcriptional repression experienced by human cells upon exposure to hyperosmotic stress and shed light on a previously unappreciated role of the Integrator complex in transcriptional regulation.

Insights into the Biogenesis of Stress-Induced Readthrough Transcripts

A Dissertation

Presented to the Faculty of the Graduate School

of

Yale University

In Candidacy for the Degree of

Doctor of Philosophy

By

Nicolle A. Rosa Mercado

Dissertation Director: Joan A. Steitz

June 2021

© 2021 by Nicolle Alexandra Rosa Mercado

All rights reserved.

TABLE OF CONTENTS

Abstract.....	i
Table of contents.....	v
List of figures.....	viii
List of abbreviations.....	x
Acknowledgements.....	xi
Aim and scope of dissertation.....	xiii
Chapter 1: Who let the DoGs out?: Insights into the biogenesis of stress-induced readthrough transcripts.....	1
Summary.....	1
Cellular stress induces readthrough transcription.....	2
Genomic and chromatin features of DoGs and their parent genes.....	5
Transcriptional landscapes accompanying DoG induction.....	8
Splicing and termination as interconnected processes.....	11
Regulation of transcription termination factors upon environmental stress contributes to DoG biogenesis.....	12
Factors contributing to DoG induction in disease.....	16
Concluding remarks.....	17
Box 1: The serendipitous discovery of DoGs.....	19
Box 2: Chromatin architecture and DoG production.....	20
Box 3: Canonical roles of termination factors important for DoG production.....	21
Glossary.....	23
Chapter 2: Hyperosmotic stress induces readthrough transcription despite widespread transcriptional repression.....	25
Summary.....	25
Introduction.....	26
Results.....	28
Hyperosmotic stress causes widespread transcriptional repression.....	28

Stress-induced readthrough transcripts arise independent of gene transcription levels.....	31
Clean DoG-producing genes are functionally enriched for transcriptional repression.....	33
Pol II is redistributed along the genome after hyperosmotic stress.....	35
Discussion.....	38
Chapter 3: Hyperosmotic stress alters the RNA Polymerase II interactome.....	42
Summary.....	42
Introduction.....	43
Results.....	45
The Pol II interactome is altered by hyperosmotic stress.....	45
Hyperosmotic stress decreases the occupancy of Integrator subunits on DNA.....	48
Depletion of Integrator endonuclease leads to DoG production.....	51
Discussion.....	56
Chapter 4: Materials and methods.....	60
Data and code availability.....	60
Experimental model and subject details.....	60
Method details.....	60
RNA preparations and RT-qPCRs.....	60
High throughput sequencing.....	61
TT-TimeLapse-seq (In collaboration with Joshua Zimmer).....	61
ChIP-seq.....	61
Bioinformatics.....	62
TT-TimeLapse-seq (In collaboration with Joshua Zimmer).....	62
ChIP-seq.....	63
Cell fractionation and immunoprecipitations.....	63
Proteomics (In collaboration with Dr. Maria Apostolidi).....	64
siRNA knockdowns.....	65

Western blots.....	65
Quantification and statistical analysis.....	66
Materials table.....	66
Oligos used in this study.....	68
Chapter 5: Discussion and future directions.....	72
Summary.....	72
Discovery of additional factors that contribute to DoG production upon stress.....	73
Assessing the temporal dynamics of transcriptional regulation and DoG production upon hyperosmotic stress and recovery.....	74
Investigating the impact of DoG transcription of read-in gene transcript levels.....	78
Outstanding questions.....	80
References.....	81

List of Figures

- Figure 1.1** Cellular localization of DoG RNAs
(page 3)
- Figure 1.2** Characteristics of DoG-producing genes.
(page 6)
- Figure 1.3** Termination factors involved in DoG production.
(page 13)
- Figure 2.1** TT-TL-seq reveals transcriptional profiles that accompany DoG induction after hyperosmotic stress.
(page 27)
- Figure 2.2** Hyperosmotic stress leads to widespread transcriptional repression.
(page 30)
- Figure 2.3** Validation of transcriptional repression induced by hyperosmotic stress at DoG-producing genes
(page 32)
- Figure 2.4** DoGs arise regardless of the transcriptional levels of their upstream genes upon hyperosmotic stress.
(page 34)
- Figure 2.5** Hyperosmotic stress causes redistribution of Pol II molecules across the genome.
(page 36)
- Figure 2.6** Anti-Pol II ChIP-seq reveals distinct occupancy patterns upon stress
(page 40)
- Figure 3.1** The Pol II interactome changes upon hyperosmotic stress.
(page 44)

Figure 3.2 Levels of certain termination factors remain unchanged after hyperosmotic stress.

(page 47)

Figure 3.3 Hyperosmotic stress leads to decreased occupancy of Integrator subunits on DNA.

(page 49)

Figure 3.4 Decreased occupancy of Integrator subunits on DNA upon hyperosmotic stress is not solely dependent on a decrease in Pol II binding

(page 50)

Figure 3.5 Depletion of Integrator nuclease subunit leads to DoG induction

(page 53)

Figure 3.6 Int11 knockdown leads to readthrough transcription

(page 55)

Figure 5.1 Cartoon depiction of DoG induction upon hyperosmotic stress or Int11 knockdown.

(page 75)

Figure 5.2 Phosphorylation of Pol II CTD residues tyrosine-1 and serine-7 does not decrease after hyperosmotic stress.

(page 77)

Figure 5.3 Dynamics of transcriptional regulation and DoG induction at DoG-producing genes.

(page 79)

List of Abbreviations

lncRNA	long noncoding RNA
DoG	downstream-of-gene transcript
HSV-1	Herpes simplex virus 1
Pol II	RNA Polymerase II
CPA	cleavage and polyadenylation
CPSF	cleavage and polyadenylation specificity factor
TES	transcription end site
TSS	transcription start site
Int	Integrator
TT-TL-seq	transient transcriptome TimeLapse sequencing
s ⁴ U	4-thiouridine
RT-qPCR	reverse transcriptase quantitative polymerase chain reaction
FC	fold change
ChIP	chromatin immunoprecipitation
siRNA	small interfering RNA
WT	wild-type
snRNA	small nuclear RNA
CTD	C-terminal domain of RNA Polymerase II
Tyr1-P	phosphorylated tyrosine 1
Ser7-P	phosphorylated serine 7

Acknowledgements

I am grateful to my advisor, Dr. Joan A. Steitz, for her continuous support of my scientific career, which precedes my entrance to Yale as a graduate student. Being a part of her lab and working alongside her have been a great honor. I feel fortunate of having a mentor who not only believes in my scientific skills, but has also served as an inspiration and as a source of unwavering support for my efforts to improve diversity in STEM.

I thank the members of my thesis committee, Drs. Karla M. Neugebauer and Matthew D. Simon, for the invaluable advice they provided throughout the course of my dissertation research. I also thank current and past members of their laboratories for technical advice and reagents. Furthermore, I thank other professors in the Department of Biological and Biomedical Sciences who have nourished my career in different ways, including Drs. Wendy Gilbert, Anthony Koleske and Daniel Colón-Ramos.

I thank my collaborators, Joshua Zimmer, Maria Apostolidi, Matthew Simon, and Jesse Rinehart, for their support and contributions to the research presented in Chapters 2, 3 and 4 of this dissertation.

I am grateful to current and former members of the Steitz lab, including Kazimierz Tycowski, Annsea Park, Johanna Withers, Mingyi Xie, Vanessa Mondol, Paulina Pawlica, Seyed Torabi, Minju Ha, Suzanne DeGregorio, Therese Yario, Mei-Di Shu, Angela Miccinello, Anna Vilborg, Andrei Alexandrov and Salehe Ghasempur, for their advice on experimental design and for the many discussions brought up during subgroups and lab meetings that allowed me to materialize the science presented in this dissertation.

I would also like to thank my classmates from the University of Puerto Rico at Cayey who have reached out throughout the course of my graduate studies offering advice,

support and validation. I am grateful for the community of Puerto Rican students, postdoctoral researchers and faculty members at Yale who have made New Haven feel closer to home.

Finally, I would like to thank my mother, Carmen Mercado, my husband, Yesse'El Gutiérrez, my sister, Zoralis Rosa, my grandparents, Zoraida Burgos and Ernesto Mercado, Irma Vega and other members of my family for their help, love and support throughout the challenging endeavor of completing graduate school while being away from home.

Aim and Scope of Dissertation

Cellular stress results in readthrough transcription, leading to the synthesis of thousands of RNAs known as downstream-of-gene containing transcripts (DoGs). The mechanisms that contribute to the biogenesis of these long noncoding RNAs upon environmental stress have remained elusive. The objective of this dissertation was to identify factors that contribute to DoG production in human cells exposed to hyperosmotic stress. Therefore, I designed experiments to investigate the nascent transcription profiles as well as the interactome of RNA Polymerase (Pol) II in human cell lines treated with concentrations of potassium chloride that ensure the production of DoGs.

Chapter 1 serves as an introduction to DoGs, the chromatin landscapes in DoG-producing regions and the factors involved in DoG synthesis induced by different types of cellular stress or disease. This chapter will be submitted as a review article and, therefore, includes findings described in subsequent chapters of this dissertation.

Chapter 2 describes the transcriptional landscapes that accompany DoG induction upon hyperosmotic stress. Here, I report that hyperosmotic stress leads to the transcriptional repression of many genes, but that a gene's ability to produce DoGs is independent of its transcriptional regulation. Furthermore, I present anti-Pol II ChIP-seq experiments that confirm the aforementioned nascent RNA sequencing results and demonstrate a redistribution of Pol II molecules into DoG regions upon KCl treatment of HEK-293T cells.

Chapter 3 focuses on experiments that investigate the protein interactions experienced by RNA Pol II upon hyperosmotic stress. These data revealed that hyperosmotic stress alters the Pol II interactome leading to a decrease in the occupancy

of Integrator complex subunits on chromatin. Furthermore, nascent RNA sequencing analyzing cells depleted of the catalytic subunit of this complex reveals DoG induction. The identity of genes that produce readthrough transcripts upon knockdown of this protein partially overlaps the identity of genes that produce DoGs upon hyperosmotic stress. Therefore, I conclude that the Integrator complex may contribute to DoG biogenesis in the context of hyperosmotic stress.

Chapter 4 describes the technical details and methods used to perform the experiments described in this dissertation and provides a comprehensive list of the materials used.

Chapter 5 presents preliminary results regarding the dynamics of transcriptional regulation experienced by DoG-producing genes upon hyperosmotic stress and recovery. ChIP-seq experiments investigating the occupancy of Pol II molecules phosphorylated on residues that are thought to mediate Pol II interactions with the Integrator complex are also discussed. Finally, this chapter provides perspective on future experiments that should be performed in order to continue expanding our understanding of the biogenesis of stress-induced readthrough transcripts in human cells.

Chapter 1

Who let the DoGs out?: Insights into the biogenesis of stress-induced readthrough transcripts

Summary:

Readthrough transcription has recently emerged as a hallmark of the cellular stress response. It results in the production of thousands of long noncoding RNAs known as downstream-of-gene containing (DoG) transcripts. DoGs arise from around ten percent of human protein-coding genes and are retained in the nucleus. These transcripts are produced minutes after cell exposure to stress and are detected hours after stress removal. However, their biogenesis and the role(s) that these RNAs or their production play in the cellular stress response are incompletely understood. Here, I discuss recent findings that implicate several host and viral proteins in the mechanisms underlying DoG production as well as the transcriptional landscapes that accompany DoG induction under different stress conditions.

Cellular stress induces readthrough transcription

Cellular stress causes a variety of changes in the molecular organization of cells, which affect cellular structure and molecular interactions. Examples include the aggregation of filaments and other proteins upon heat shock (reviewed in Richter et al. 2010) as well as nuclear shrinkage and chromatin collapse upon hyperosmotic stress (Finan and Guilak. 2010; Finan et al. 2011). Perturbations in the transcriptional profiles of cells due to stress have also been documented (De Nadal et al. 2011; Himanen and Sistonen. 2019; Vihervaara et al. 2018; Shalgi et al. 2014; Mahat et al. 2016; Rosa-Mercado et al. 2021). Recently, several groups have found that cellular stress, including hyperosmotic stress, heat shock (Vilborg et al. 2015), oxidative stress (Vilborg et al. 2017), renal cancer (Grosso et al. 2015), hypoxia (Wiesel et al. 2018), and certain viral infections (Rutkowski et al. 2015; Bauer et al. 2018) inhibit the 3'-end cleavage of nascent RNAs arising from certain genes, leading to **readthrough transcription (See Glossary)**. This stress-induced readthrough transcription results in the production of downstream-of-gene containing (DoG) transcripts (reviewed in Vilborg and Steitz. 2017; **Box 1**).

DoGs arise from around ten percent of human protein-coding genes and are continuous with their upstream mRNAs (Vilborg et al. 2015). Therefore, the production of these RNAs does not require *de novo* transcription initiation from an independent transcription start site (TSS). DoGs are defined as having a minimal length of five kilobases beyond the annotated transcription end site (TES) of their parent gene, but can

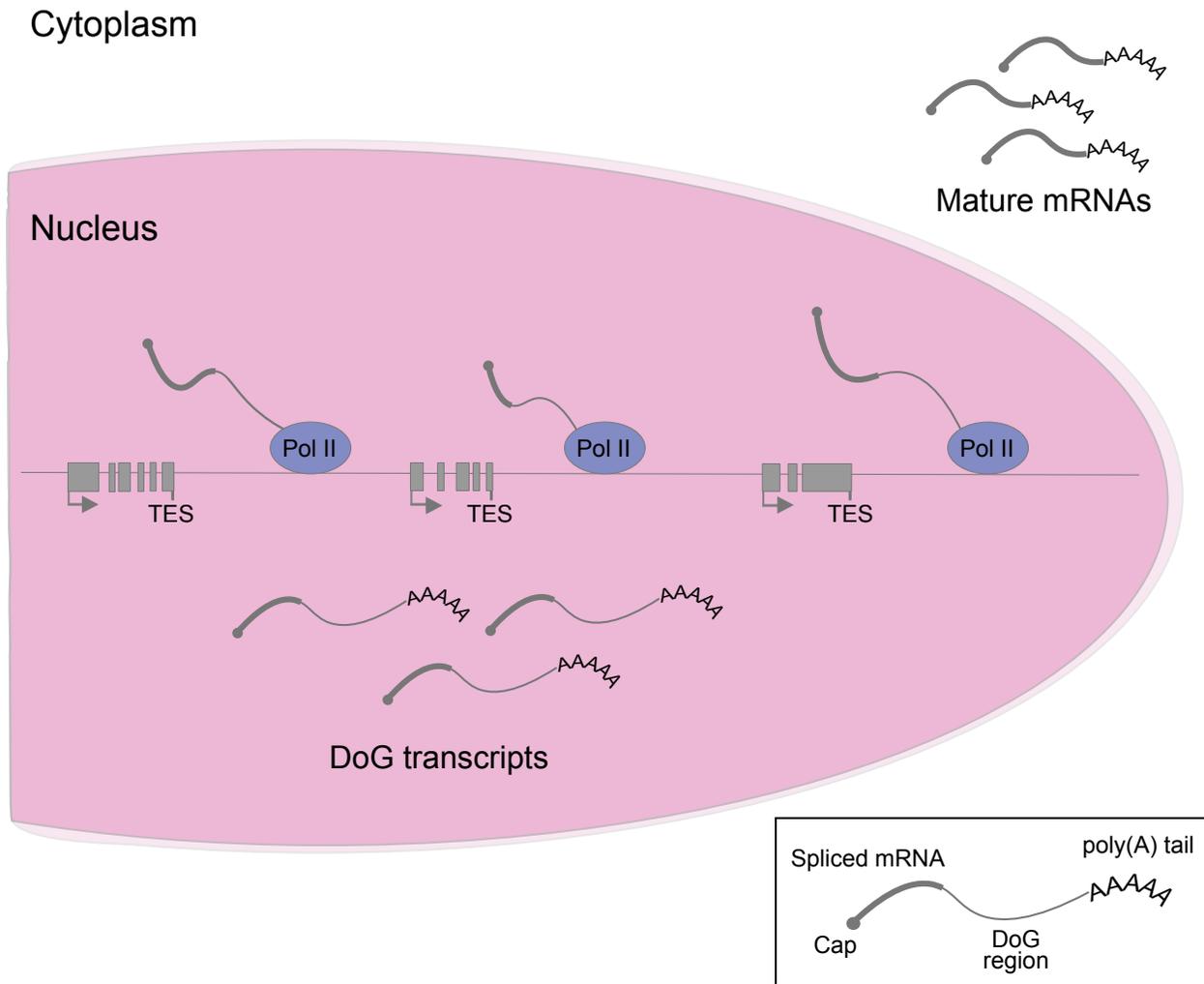


Figure 1.1: Cellular localization of DoG RNAs. DoG transcripts, which are continuous with their upstream mRNA, are retained in the nucleus. After synthesis, many DoGs remain close to their sites of transcription, while others are released into the nucleoplasm (Vilborg et al. 2015, 2017; Hennig et al. 2018). DoGs have been detected at similar levels in both polyadenylated and unpolyadenylated forms (Vilborg et al. 2015).

reach lengths of up to two hundred kilobases in mammalian cells. The fraction of Pol II molecules that read through a DoG-producing gene upon stress remains unknown. The identity of **DoG-producing genes** shows significant overlap within a given cell line exposed to different types of stress (Vilborg et al. 2017, Hennig et al. 2018). Yet, it is unclear whether the same set of genes produces DoGs across different cell types.

Transcription units with other genes within 50 kb of their 3'-ends are more likely to experience readthrough transcription than those in gene-poor chromosomal regions (Vilborg et al. 2017). Due to their lengths, DoGs often overlap with downstream transcription units, which are referred to as "**read-in genes**" (Rutkowski et al. 2015; Grosso et al. 2015; Roth et al. 2020). The impact of DoG transcription on the expression of read-in genes is currently unclear (further discussed below).

DoGs are induced minutes after cells are exposed to environmental stressors, such as heat shock or hyperosmotic stress. However, the time required for DoGs to return to basal levels after stress removal ranges from several hours to a day, depending on the type of stress (Vilborg et al. 2017). Moreover, previous work has demonstrated that the half-life of DoGs is about an hour (Vilborg et al. 2015), similar to the median half-life of mRNAs (Schwalb et al. 2016). This reveals that DoG transcripts are not quickly degraded after production.

DoG transcripts are retained in the nucleus and many remain close to their site of transcription, while others are detected as nucleoplasmic puncta (Vilborg et al. 2015; Vilborg et al. 2017; Hennig et al. 2018; Figure 1.1). One possibility is that DoGs or their production are important for supporting chromatin structure throughout the cellular stress response (Vilborg et al. 2015; Vilborg and Steitz. 2017). However, experiments that

directly measure the impact of DoG RNAs and their transcription will be necessary to gain further insight into the potential effects of these RNAs on chromatin.

Genomic and chromatin features of DoGs and their parent genes

In an effort to understand why DoGs arise from only ten percent of protein-coding genes upon hyperosmotic stress, several groups have investigated whether DoG-producing genes have signature chromatin features or sequences that are less prevalent in genes that fail to produce DoGs (**non-DoG genes**). Interestingly, genes that produce readthrough transcripts in the context of HSV-1 infection tend to have weaker polyadenylation signals than non-DoG genes (Rutkowski et al. 2015). Moreover, previous reports show that DoG regions themselves are also slightly depleted of the canonical poly(A) signal, AATAAA, and lack other motifs important for efficient transcription termination (Vilborg et al. 2015; Vilborg et al. 2017; Figure 1.2). These findings suggest that DoG-producing genes are predisposed to readthrough transcription.

Recent studies reveal that genes that produce DoGs after cells have been treated with the splicing inhibitor, PladB, have lower levels of termination-associated R-loops than non-DoG genes (Castillo-Guzman et al. 2020). These terminal R-loops have been previously shown to be important for efficient transcription termination (Skourti-Stathaki et al. 2011). Therefore, the authors propose that the observed depletion of terminal R-loops could explain why DoG-producing genes are prone to readthrough transcription (Castillo-Guzman et al. 2020). Investigating the prevalence of terminal R-loops in the 3'-end of genes that produce DoGs upon different types of cellular stress will reveal if the observed depletion is a general characteristic of all DoG-producing genes. Such analyses will also provide more information regarding the mechanistic relevance of R-loops to DoG

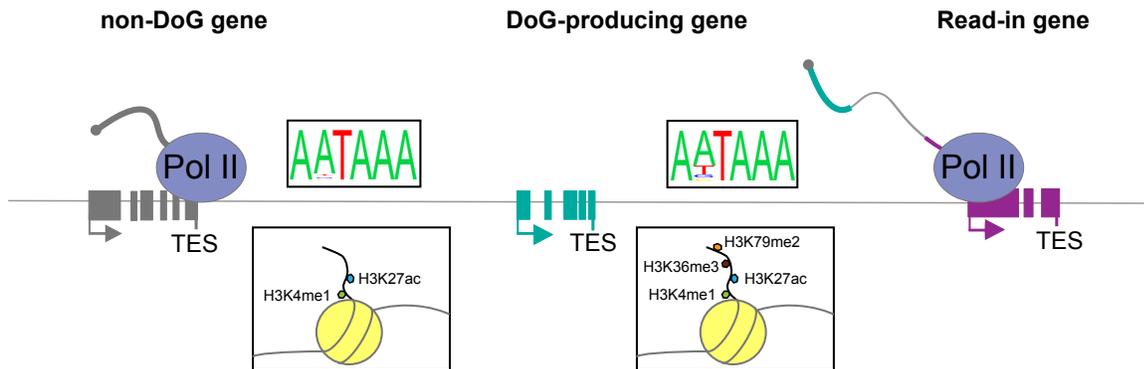


Figure 1.2: Characteristics of DoG-producing genes. According to previous estimates, a canonical poly(A) signal is found within 50 base pairs of the 3'-end of 79% of non-DoG genes and 61% of cellular genes that produce DoGs upon HSV-1 infection (Rutkowski et al. 2015). Similarly, comparing the frequency of appearance of the canonical poly(A) signal in the sense versus the antisense strand across the region 5 kb downstream of the annotated transcription end site (TES) of a DoG-producing gene induced by hyperosmotic stress revealed that this sequence is less prevalent in DoG-producing genes compared to non-DoG genes (ratio of 0.8 at DoG-producing genes versus 1.1 in non-DoG genes) (Vilborg et al. 2015). The estimated prevalence of the canonical poly(A) signal, AATAAA and of the weaker ATTAAA are depicted in the figure. Additionally, DoG regions are slightly enriched for histone marks that favor transcription elongation throughout the first 5 kb after their annotated TES. These regions are also enriched for histone marks typically found in intergenic regions, such as H3K4me1 and H3K27ac (Vilborg et al. 2017). Transcription of DoGs often leads to read-in transcription of downstream genes (Rutkowski et al. 2015; Roth et al. 2020; Rosa-Mercado et al. 2021).

biogenesis.

Studies assessing the chromatin environment of stress-induced DoG-producing genes prior to stress (homeostasis) have uncovered additional characteristic features that promote elongation beyond the polyadenylation signal of these transcription units. Histone marks contribute to the progression of transcription by recruiting protein factors that mediate various steps of the process or by favoring certain chromatin conformations (Bannister and Kouzarides. 2011; Gates et al. 2017). Previous work demonstrates that the 3'-ends and downstream regions of genes that produce DoGs common to different types of cellular stress (**pan-stress DoGs**) are slightly enriched for H3K79me2 and H3K36me3, histone marks, which favor transcription elongation (Vilborg et al. 2017; Huang and Zhu. 2018). Enrichment of H3K4me1 and H3K27ac was also observed in pan-stress DoG regions. However, analyses evaluating the distribution of these two histone marks also revealed an enrichment across all intergenic regions, consistent with their presence in enhancers (Woo et al. 2012; Heintzman et al. 2007; Creighton et al. 2010; Figure 1.2). The enrichment of H3K36me3 in regions downstream of genes was confirmed by the Friedel and Dölken groups in the context of early HSV-1 infection (Hennig et al. 2018). Yet, cells exposed to prolonged periods of HSV-1 infection produced different results, demonstrating that instead there exists a negative correlation between general (H3K4me3) and repressive (H3K27me3) promoter marks in DoG-producing chromatin regions (Hennig et al. 2018). Overall, how chromatin marks change at DoG-producing genes in response to different stressors is yet to be assessed (Hennig et al. 2018; Vilborg et al. 2017). So far, studies have only begun to uncover mechanisms that

favor transcription downstream of the annotated transcription end sites of a subset of genes, but many additional factors likely contribute (further discussed below).

Chromatin organization is characterized in part by transcriptional activity (Lieberman-Aiden et al. 2009; Sexton et al. 2012; Ulianov et al. 2016; **Box 2**). Moreover, regions of active transcription generally exhibit hallmarks of open chromatin conformation, whereas regions that are transcriptionally silent have closed chromatin features (Li et al. 2004). In untreated cells, open chromatin features are prevalent near the promoters and in downstream regions of pan-stress DoG-producing genes as measured by DNase-seq and ATAC-seq experiments (Vilborg et al. 2017). However, the prevalence of open chromatin was not observed when other DoG-regions that are specifically induced upon hyperosmotic stress or heat shock were included in analyses of unstressed cells (Hennig et al. 2018). Consistently, ATAC-seq experiments performed on cells exposed to hyperosmotic stress or heat shock for one or two hours did not show increased chromatin accessibility in DoG regions upon stress. On the other hand, an increase in chromatin accessibility was observed across readthrough regions after HSV-1 infection (Hennig et al. 2018). Such discrepancies further emphasize the need for a comprehensive mechanistic understanding of DoG induction by specific stress conditions versus pan-stress DoG induction.

Transcriptional landscapes accompanying DoG induction

Multiple cellular stress conditions that lead to the production of DoGs also induce other changes that impact the transcriptional landscape of cells. Recently, **TT-TimeLapse (TT-TL) sequencing** (Schofield et al. 2018) was combined with the use of spike-in RNA controls to investigate the nascent transcription profiles that accompany DoG production

after hyperosmotic stress. Analysis of a subset of genes that were free from invasive transcription originating in neighboring DoG-producing genes (**clean genes**), revealed that hyperosmotic stress induces widespread transcriptional repression (Rosa-Mercado et al. 2021). Accordingly, recent anti-Pol II ChIP-seq experiments also demonstrate a decrease in Pol II occupancy across gene bodies after hyperosmotic stress, which is accompanied by a redistribution of Pol II molecules into downstream-of-gene regions (Amat et al. 2019; Rosa-Mercado et al. 2021). The decreased occupancy of Pol II molecules across gene bodies and the consequent transcriptional repression induced by hyperosmotic stress were explained at least in part by an increase in termination near transcriptional start sites, which prevents Pol II from progressing into the elongation phase (Zimmer et al. 2021). Other reports have demonstrated transcriptional repression in the context of viral infection or heat shock, which was similarly accompanied by a shift of Pol II molecules into downstream regions (Birkenheuer et al. 2018; Mahat et al. 2016; Bauer et al. 2018; Cardiello et al. 2018). Upon heat shock, inhibition of Pol II pause release was shown to contribute to transcriptional repression (Mahat et al. 2016). These results highlight common features of the chromatin landscape accompanying DoG transcription and demonstrate the impact of DoG production on Pol II distribution.

Measurements of Pol II occupancy and nuclear RNA levels have shown that DoGs arise from genes that fall into different categories of transcriptional regulation upon stress (Vilborg et al. 2017; Cardiello et al. 2018). Consistent with these observations, recent TT-TL-seq experiments demonstrate that DoGs can arise independent of the transcriptional response of their parent gene upon hyperosmotic stress. Namely, a similar fraction of

genes that are activated, unchanged or repressed are able to produce DoGs (Rosa-Mercado et al. 2021).

It has become evident that DoG production extensively impacts the accompanying transcriptional landscape. For instance, upon HSV-1 infection or hyperosmotic stress, a significant fraction of expressed genes experiences read-in transcription, which may affect the expression of downstream genes through transcriptional interference (Rutkowski et al. 2015; Rosa-Mercado et al. 2021; Mazo et al. 2007). From a technical perspective, such read-in transcription complicates differential expression analysis, making it seem that many genes are transcriptionally activated by stress (Cardiello et al. 2018; Roth et al. 2020). Furthermore, DoG RNAs could also serve as antisense transcripts whose production may reflect negative effects on the transcription of read-in genes on the opposite strand (Vilborg et al. 2017; Muniz et al. 2017), perhaps contributing to the maintenance of transcriptional repression upon stress (Pelechano and Steinmetz. 2013).

Previous functional assessments of DoG-producing genes have been complicated by the fact that read-in transcription often causes DoGs to be assigned to incorrect parent genes (Rosa-Mercado et al. 2021). Therefore, Hennig et al. (2018) investigated functional aspects of DoG-producing genes by taking a targeted approach. They observed that the *IRF1* gene, which encodes a transcription factor involved in regulating the host immune response (Forero et al. 2019), produces DoGs upon viral infection, heat shock or hyperosmotic stress. By taking advantage of their previously-published ribosome profiling experiments, which assessed the impact of HSV-1 infection on translation (Rutkowski et al. 2015), the authors demonstrated that translation of the *IRF1* mRNA decreases upon

infection. One possible explanation for this observation might be that DoGs play a role in dampening the host immune response in the context of viral infection by retaining unprocessed *IRF1* transcripts in the nucleus (Hennig et al. 2018). Conversely, recent gene ontology analyses reveal that clean DoG-producing genes induced upon hyperosmotic stress are functionally enriched for terms related to transcriptional repression. This finding supports the idea that DoGs might serve as a reservoir of unprocessed transcripts that could be used over time to perpetuate a transcriptionally-repressed state upon prolonged exposure to stress (Rosa-Mercado et al. 2021). Whether DoGs represent a strategy that cells employ to survive extended periods of stress or a response orchestrated by viruses to promote their own replication is yet to be determined.

Splicing and termination as interconnected processes

Co-transcriptional processing of pre-mRNAs demands interconnected steps. Interestingly, studies in yeast and mammalian cells have revealed that unspliced pre-mRNAs are not cleaved at their 3'-ends and, therefore, exist as readthrough transcripts (Herzel et al. 2018; Alpert et al. 2020; Reimer et al. 2021). In budding yeasts, readthrough transcription is induced upon depletion of Nab2, which binds polyadenylated RNAs and was thought to be involved in splicing (Soucek et al. 2016). Long read sequencing experiments revealed that splicing was only disrupted for transcripts arising from read-in transcription events, suggesting that the chimeric transcripts produced from intrusive transcription are not optimal splicing substrates (Alpert et al. 2020). Furthermore, experiments performed in differentiating mouse erythroblasts demonstrate that readthrough transcripts arise from genes whose pre-mRNAs experience inefficient co-transcriptional splicing. Targeted long-read sequencing of transcripts arising from the

beta-globin gene in these murine erythroleukemia cells further revealed that introducing a cryptic splice site that increases splicing efficiency also increases 3'-end cleavage, indicating a functional relationship between these processes (Reimer et al. 2021). Readthrough transcription of stress-responsive genes has also been observed in the presence of the splicing inhibitor, PladB (Castillo-Guzman et al. 2020). Consistent with previous observations, genes producing mRNAs with retained terminal introns are more likely to read through than other genes, further suggesting that splicing disruptions interfere with 3'-end cleavage and transcription termination for at least a subset of genes.

Disrupted splicing patterns have also been reported in several conditions that induce DoG production, including heat shock (Shalgi et al. 2014) and HSV-1 infection. Upon HSV-1 infection, splicing disruptions were predominantly observed in read-in genes (Rutkowski et al. 2015). Further examination of the impact of cellular stress on splicing revealed that read-in transcription results in intergenic splicing, leading to the formation of chimeric exon-exon junctions (Hennig et al. 2018). This phenomenon was more evident after HSV-1 infection than after heat shock or hyperosmotic stress. These observations further suggest that the transcriptional machinery responsible for this invasive transcription is not always splicing-competent. More experiments that investigate the splicing status of DoG RNAs are necessary to fully understand the correlations observed in cells exposed to stress.

Regulation of transcription termination factors upon environmental stress contributes to DoG biogenesis

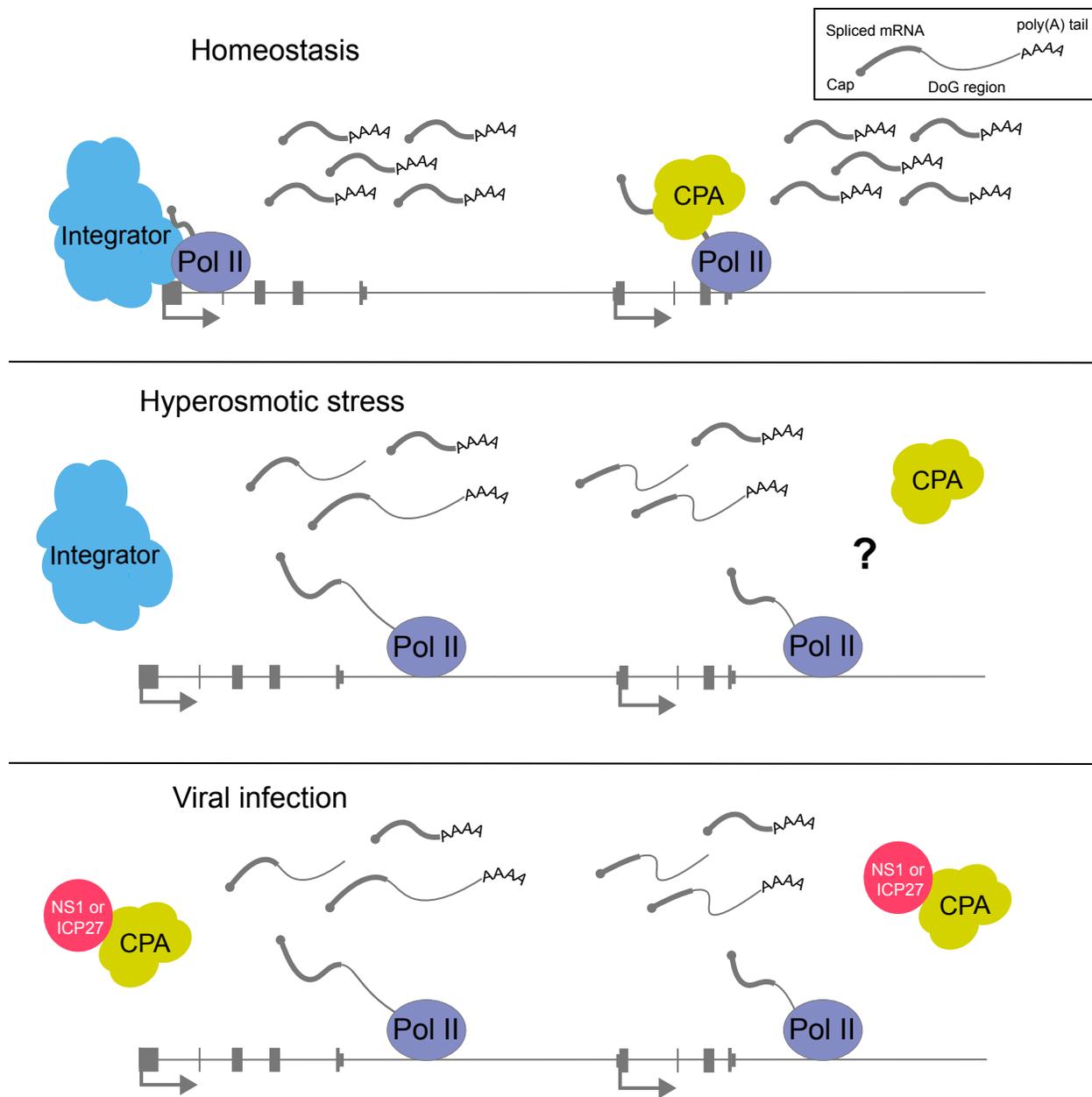


Figure 1.3: Termination factors involved in DoG production. Under homeostasis, components of the cleavage and polyadenylation machinery mediate the 3'-end formation of protein-coding gene transcripts (Neve et al. 2017), while the Integrator complex regulates RNA output near transcription start sites (Elrod et al. 2019; Becketdorff et al. 2020). Upon hyperosmotic stress, the interaction between Pol II and Integrator is disrupted and DoG-production is observed (Rosa-Mercado et al. 2021). Dysregulated occupancy of CPA components on DNA has been reported after hyperosmotic stress (Jalihal et al. 2020). However, whether this deficit directly affects DoG-producing genes remains unclear. The uncertain (yet likely) loss of CPA factors at DoG-producing genes is depicted in the figure with a question mark. After Influenza A or HSV-1 infection, viral proteins NS1 and ICP27, respectively, sequester subunits of the CPA machinery leading to DoG biogenesis (Nemeroff et al. 1998; Wang et al. 2020).

DoGs arise as a consequence of decreased transcription termination (Vilborg et al. 2015; **Box 3**). Recent studies have shed light on how cellular stress affects the occupancy of different termination factors on chromatin. ChIP analyses of several components of the CPA machinery, including CPSF73, CPSF6 and CPSF1, show that their occupancy at the TES of certain genes decreases after hyperosmotic stress or heat shock (Jalihal et al. 2020; Cardiello et al. 2018). Moreover, CPSF6 was shown to be sequestered into phase-separated droplets upon hyperosmotic stress (Jalihal et al. 2020). Consistent with the role of CPA factors in catalyzing the 3'-end formation of protein-coding genes, depletion of CPSF73 leads to readthrough transcription (Vilborg et al. 2015; Eaton et al. 2020). Cardiello et al. (2018) and Jalihal et al. (2020) suggest the direct involvement of CPA factors in DoG biogenesis upon stress. However, a thorough analysis assessing the differential binding of these factors to DoG-producing versus non-DoG genes in stressed cells will be necessary to confirm this notion (Figure 1.3).

To identify termination factors that contribute to DoG induction upon stress, mass spectrometry analyses of anti-Pol II immunoprecipitates obtained from untreated and KCl-treated cells were performed. Results obtained from these experiments demonstrate that hyperosmotic stress alters the Pol II interactome (Rosa-Mercado et al. 2021). Surprisingly, mass spectrometry and western blot analyses revealed the presence of certain CPA factors and of Xrn2, the nuclease that mediates the final step of transcription termination (Fong et al. 2015), in the anti-Pol II immunoprecipitates obtained from chromatin fractions of untreated and KCl-treated cells. Conversely, interactions between Pol II and subunits of the Integrator complex decreased upon stress (Rosa-Mercado et al. 2021). Furthermore, ChIP-seq experiments using antibodies against two different

subunits of the Integrator complex showed a genome-wide loss of occupancy at both DoG-producing genes and non-DoG genes. Knockdown of the catalytical subunit of Integrator, Int11, led to the induction of hundreds of readthrough transcripts. Comparison of genes that produce readthrough transcripts after Int11 knockdown to clean genes that produce DoGs upon hyperosmotic stress showed an overlap of up to 25%, suggesting that the Integrator complex partially contributes to DoG biogenesis in this context (Rosa-Mercado et al. 2021). However, it remains unclear whether the Integrator complex plays a direct role in DoG induction. Further assessment of the transcriptional machinery present near the TESs of DoG-producing genes will reveal additional features that make these genes prone to readthrough transcription upon stress.

It is possible that cells employ stress-specific pathways in order to orchestrate DoG production. This idea is supported by the potential role of the temperature-responsive transcription factor, HSF1, in regulating DoG biogenesis upon heat shock. Analyses of heat-shocked cells demonstrate that the level of readthrough transcription downstream of a subset of genes is dictated by HSF1 binding to their promoters (Vilborg et al. 2017). Furthermore, initial studies suggested that upon hyperosmotic stress, but not heat shock, DoG induction was mediated by calcium signaling through the IP3 receptor (Vilborg et al. 2015; Vilborg et al. 2017; Vilborg and Steitz. 2017). However, it was later discovered that the decrease in DoG levels observed after inhibiting calcium signaling was caused by a general decrease in transcription (Hennig et al. 2018). Therefore, additional studies addressing the stress-specific pathways that regulate termination factors involved in DoG biogenesis are necessary to understand the signaling strategies that cells use to initiate this rapid response to stress.

Factors contributing to DoG induction in disease

DoG production has been observed in several disease contexts including renal cancer, HSV-1 infection and Influenza A infection. Viruses enter cells equipped to hijack host processes in order to prioritize their own replication (Withers et al. 2018). It has been proposed that the production of DoGs could facilitate viral replication by decreasing the number of host mRNAs exported to the cytoplasm (Nemeroff et al. 1998). The Influenza A protein NS1 was observed to sequester CPSF30, leading to the production of readthrough transcripts from host genes upon infection (Nemeroff et al. 1998; Figure 1.3). However, later studies found that the induction of DoGs during Influenza A infection is independent of interactions between NS1 and CPSF30 (Bauer et al. 2018). It is possible that NS1 is sufficient to inhibit termination, but other factors that contribute to viral infection may ensure readthrough transcription of host genes in its absence. Similarly, the HSV-1 protein, ICP27, mediates viral disruption of host transcription termination by interfering with the assembly of host CPA complexes (Wang et al. 2020; Figure 1.3). These findings showcase the purposeful induction of readthrough transcripts upon viral infection.

Readthrough transcription has also been observed in tumor samples obtained from patients suffering from clear cell renal cell carcinoma. Here, high levels of readthrough transcription inversely correlate with survival (Grosso et al. 2015). One of the most frequently mutated genes in patients with this malignancy is *SETD2*, which encodes a histone modifying enzyme that mediates the trimethylation of H3K36 (Duns et al. 2010). Interestingly, as discussed above, H3K36me3 has been found to be enriched in DoG regions (Figure 1.2). Importantly, expression of wild-type *SETD2* in mutant cell lines was

sufficient to revert the defect in transcription termination at some DoG-producing genes (Grosso et al. 2015).

DoGs have also been detected in senescence, where cells are not proliferative, in contrast to cancerous cells. Although beneficial in certain contexts, senescence has also been associated with endocrine disease and ageing (Khosla et al. 2020). In senescent cells, DoGs are thought to play a key role in controlling gene expression by acting as antisense transcripts. Compared to control cells, Muniz et al. (2017) observed that senescent cells exhibited decreased occupancy of H2A.Z, an alternative histone involved in transcriptional regulation (Giaimo et al. 2019). The authors propose that the presence of the histone variant, H2A.Z, represses readthrough transcription under normal cellular conditions. However, they argue that other factors are likely involved in the induction of readthrough transcripts upon senescence because knockdown of H2A.Z leads only to partial DoG induction. These studies further highlight the need for a better understanding of how important chromatin modifiers and histone marks respond to disease and other stress conditions.

Concluding remarks

Stress-induced readthrough transcription has been shown to be a prevalent marker of the cellular stress response upon exposure to certain environmental stressors or disease. The transcripts that arise have been thoroughly characterized over the last few years. Yet, much work is still required to gain insights into their functions, as well as a complete picture of the mechanisms responsible for their production.

Recent findings have expanded our understanding of the biogenesis of stress-induced readthrough transcripts. The data presented in the manuscripts reviewed here suggest

the involvement of termination factors, including the Integrator complex (Rosa-Mercado et al. 2021) and components of the CPA machinery (Cardiello et al. 2018; Jalihal et al. 2020), as well as histone modifying enzymes in modulating the biogenesis of these stress-induced long noncoding RNAs. The role of subunits of the CPA complex in readthrough transcription upon viral infection has been thoroughly characterized (Nemeroff et al. 1998; Bauer et al. 2018; Wang et al. 2020). However, how the CPA complex is regulated at DoG-producing genes upon environmental stress is less clear. Studies focused on cancerous or senescent cells have described readthrough transcription in both these contexts and suggest that DoG production might be regulated by characteristic changes in histones and histone marks (Grosso et al. 2015; Muniz et al. 2017). It is possible that cells undergoing long lasting changes, such as cancer cells, rely on histone modifiers to regulate DoG production, whereas conditions that require a rapid response mediate DoG formation through the differential occupancy of termination factors. Additionally, studies of co-transcriptional splicing have revealed that transcripts with retained introns are prone to failed 3'-end cleavage (Herzel et al. 2018; Alpert et al. 2020; Reimer et al. 2020). The interplay between these mechanisms is yet to be investigated and promises to provide a more comprehensive picture of how termination is regulated across different genes.

The association of DoG induction with disease and disease-like contexts, such as heat shock and hyperosmotic stress, poses the question of whether DoGs or their production play a protective role in cells or whether they promote disease progression. Whatever the case, mechanistic studies of DoG biogenesis reviewed here have laid the foundation for future functional studies and therapeutic applications.

Box 1: The serendipitous discovery of DoGs

DoGs were originally characterized in 2015 when three groups independently published observations of widespread readthrough transcription in different contexts: including hyperosmotic stress (Vilborg et al. 2015), HSV-1 infection (Rutkowski et al. 2015) and clear cell renal cell carcinoma (Grosso et al. 2015). These groups had set out to study different aspects of transcriptional regulation in cells challenged with different stressors.

Vilborg et al. (2015) aimed to characterize a long noncoding RNA (lncRNA) that had been shown to correlate with unfavorable outcomes in neuroblastoma patients. To study the transcriptional regulation of this putative lncRNA, human neuroblastoma cells were exposed to hyperosmotic stress, considered at the time to be a transcriptional activator of neuronal cells (Kim et al. 2010; Greer and Greenberg. 2008). Assessment of nuclear RNA-seq data led our lab to the realization that the lncRNA of interest was in fact continuous with the transcript arising from the upstream gene, *CXXC4*, and was, therefore, a product of readthrough transcription. Hence, doCXXC4 (downstream of CXXC4) was the first transcript to be baptized as a DoG. Additional analyses of nuclear RNA-seq data revealed the induction of DoGs from thousands of protein-coding genes upon KCl treatment.

Concurrently, the Dölken lab set out to understand the impact of HSV-1 infection on the transcription, processing and translation of host mRNAs by combining ribosome profiling with sequencing experiments targeting recently synthesized RNAs. To study recently produced transcripts, they sequenced RNAs labeled with the uridine analogue 4-thiouridine (s⁴U) during a one hour incubation period at various timepoints after infection.

This approach revealed defects in the transcription termination of many host genes, resulting in the production of DoGs (Rutkowski et al. 2015).

Meanwhile, Grosso et al. (2015) were interested in understanding the RNA populations in samples obtained from clear cell renal cell carcinoma tumors. By comparing RNA-seq data from these tumors with data obtained from normal matched tissues, they were able to detect accumulation of readthrough transcripts in samples obtained from renal cancer patients.

The serendipitous observations of these three labs created a new field of research that promises to uncover additional layers of complexity in gene regulation, transcription termination and the interconnections between different steps of transcription.

Box 2: Chromatin topology and DoG production

Chromatin organization is divided into several layers that facilitate important interactions between enhancers, promoters and their target genes. These layers include topologically associated domains (TADs), which span tens to hundreds of kilobases and have defined boundaries established by the binding of architectural proteins such as CTCF and cohesin (Beagan and Phillips-Cremens. 2020). An additional layer of chromatin organization are A/B compartments, defined based on transcriptional activity and the presence of activating or repressive histone marks. The establishment and maintenance of chromatin architecture are critical for programs of gene expression and are altered in different disease states.

Several groups have published genome-wide assessments of chromatin reorganization upon stress. A recent study demonstrated that cells preserve A/B

compartments and TAD boundaries upon heat shock (Ray et al. 2019). In contrast, upon hyperosmotic stress or influenza A infection, there is a reorganization of chromatin compartments accompanied by a loss of CTCF and cohesin (Amat et al. 2019; Heinz et al. 2018). Specifically, after Influenza A infection, the CTCF and cohesin molecules that bind DoG regions are displaced by elongating Pol II molecules (Heinz et al. 2018). In agreement with these observations, experiments in cells exposed to hyperosmotic stress show that the redistribution of Pol II molecules into DoG regions weakens TAD boundaries, establishing a role for DoGs in the observed rearrangement of local chromatin organization (Amat et al. 2019). Moreover, stress-recovery experiments demonstrate that chromatin organization is restored one hour after removing cells from hypertonic media. Yet, exposing cells to hyperosmotic stress in combination with an inhibitor of transcription initiation suggests that some aspects of the observed chromatin reorganization are independent of new transcription (Amat et al. 2019), contradicting observations in Influenza A-infected cells (Heinz et al. 2018).

Box 3: Canonical roles of transcription termination factors important for DoG production

Transcription termination of protein-coding genes requires cleavage of the nascent mRNA after RNA Polymerase (Pol) II transcribes across the polyadenylation signal (PAS) of a gene. After the cleavage and polyadenylation (CPA) machinery cleaves the nascent transcript, Pol II continues to transcribe, producing an uncapped RNA. Removal of Pol II from the DNA template is induced when the 5' to 3' exonuclease Xrn2 crashes into Pol II as it degrades the attached uncapped transcript downstream of the polyadenylation signal (Fong et al. 2015). This model of termination is known as the torpedo model. Another

model for transcription termination, known as the allosteric model, posits that Pol II undergoes a conformational change when it transcribes through the polyadenylation signal of a gene, resulting in its dissociation from the DNA template (Zhang et al. 2015). These models have been shown to work in combination to achieve efficient transcription termination at the ends of mammalian protein-coding genes (Eaton et al. 2020). Therefore, it is not surprising that depletion of either CPSF73 or Xrn2 leads to genome-wide readthrough transcription (Fong et al. 2015; Eaton et al. 2020).

The Integrator (Int) complex plays important roles in the transcription termination of different types of noncoding RNAs. Subunits of its cleavage module, composed of Int 4, 9 and 11, mediate the 3'-end formation of pre-snRNAs (Baillat et al. 2005; Albrecht and Wagner. 2012, Albrecht et al. 2018) and of certain viral microRNAs (Xie et al. 2015). Depletion of Integrator subunits has also been found to impair transcription termination of replication-dependent histone genes, enhancer RNA genes and lncRNA genes (Skaar et al. 2015; Lai et al. 2015; Nojima et al. 2018). Furthermore, the Integrator complex plays a role in promoter-proximal Pol II pausing at protein-coding genes (Stadelmayer et al. 2014; Elrod et al. 2019) and is believed to police polymerases that are unfit to properly proceed into transcription elongation (Lykke-Andersen et al. 2021). Recently, depletion of the catalytic subunit of the Integrator complex was shown to induce hundreds of DoGs (Rosa-Mercado et al. 2021).

Glossary

- **Clean genes:** Expressed genes that do not experience invasive transcription from neighboring DoG-producing genes (Rosa-Mercado et al. 2021).
- **Cleavage and polyadenylation machinery:** Composed of four different protein complexes (CPSF, CstF, CFIm and CFII_m) that catalyze the cleavage step allowing subsequent polyadenylation of nascent pre-mRNAs from protein-coding genes in metazoans by recognizing the polyadenylation signal as well as other auxiliary motifs that facilitate the process (Neve et al. 2017).
- **DoG-producing genes:** Genes that produce readthrough transcripts of a length of five kilobases or more beyond the annotated transcription end sites (Vilborg et al. 2015).
- **H2A.Z:** Histone variant involved in regulating the transcriptional output of genes (Giaimo et al. 2019).
- **Integrator complex:** Protein complex composed of fourteen subunits that catalyzes the 3'-end formation of several noncoding RNAs (Baillat et al. 2005; Albrecht and Wagner 2012). This complex has also been implicated in promoter proximal Pol II pausing near the TSS of protein-coding genes (Stadelmayer et al. 2014; Elrod et al. 2019).
- **Long noncoding RNA:** RNA molecules longer than 200 nucleotides that do not undergo translation. Many lncRNAs have roles in maintaining chromatin structure and in regulating gene expression (Yao et al. 2019).
- **Non-DoG genes:** Genes that undergo efficient transcription termination despite exposure to cellular stress (Vilborg et al. 2015).

- **Pan-stress DoGs:** DoGs that are consistently induced by different types of cellular stress in a given cell line (Vilborg et al. 2017).
- **Readthrough transcription:** Refers to transcription extending beyond the annotated transcription end site of a gene.
- **Read-in genes:** Loci that are invaded by readthrough transcription arising from neighboring DoG-producing genes (Rutkowski et al. 2015; Roth et al 2020).
- **SETD2:** A histone methyl transferase responsible for the trimethylation of histone 3 lysine 36 (Duns et al. 2010). The gene that encodes this protein has been found to be heavily mutated in renal cancer patients (Grosso et al. 2015).
- **TT-TimeLapse sequencing:** Transient Transcriptome TimeLapse sequencing is a technique that combines short periods of metabolic labeling of RNAs using the uridine analogue, s^4U , with streptavidin-based enrichment strategies in order to isolate nascent RNAs (Duffy et al. 2015; Schwalb et al. 2016; Schofield et al. 2018). The uridine analogue in the isolated RNAs is then recoded to a cytosine analogue using TimeLapse chemistry, thereby decreasing the likelihood of contaminating reads being called as nascent RNAs.

Chapter 2

Hyperosmotic stress induces readthrough transcription despite widespread transcriptional repression

Summary

Stress-induced readthrough transcription results in the synthesis of downstream-of-gene (DoG) containing transcripts. The nascent transcription profiles that accompany DoG induction upon hyperosmotic stress had not been investigated. Therefore, the transcriptional response of DoG-producing genes was not well understood. In collaboration with Joshua Zimmer in Professor Matthew Simon's lab, I performed TT-TimeLapse sequencing experiments to assess the nascent transcription profiles during DoG induction in human cell lines. Data obtained from these experiments revealed transcriptional repression of thousands of genes upon hyperosmotic stress. Yet, DoGs are produced independent of the transcriptional level of their parent gene. ChIP-seq confirmed that stress-induced redistribution of RNA Polymerase (Pol) II correlates with the transcriptional output of genes. The work presented in this chapter clarifies the transcriptional response of human cells to hyperosmotic stress and provides evidence demonstrating that transcriptional regulation of DoG-producing genes and DoG induction are not mechanistically linked.

The work presented in this chapter has been published in *Molecular Cell* (Rosa-Mercado et al. 2021).

Introduction

Stress-induced readthrough transcripts have been previously characterized using nuclear RNA sequencing, 4sU-sequencing, mNET-seq and total RNA sequencing (Vilborg et al. 2015; Rutkowski et al. 2015; Hennig et al. 2018; Bauer et al. 2018; Grosso et al. 2015). While these techniques enable the genome-wide identification of DoGs, most do not provide quantitative information regarding the nascent transcription profiles that coincide with DoG induction.

Transient Transcriptome (TT) sequencing coupled with TimeLapse (TL) chemistry (Schwalb et al. 2016; Schofield et al. 2018) was used to characterize transcriptional profiles that accompany DoG induction upon hyperosmotic stress. This method enables reliable detection of nascent RNAs by using short pulses of metabolic labeling coupled with transcript enrichment and nucleoside recoding chemistry. Here, I report that hyperosmotic stress causes widespread transcriptional repression, from which a minority of human genes escape. Yet, DoGs are induced from genes that experience different transcriptional responses to hyperosmotic stress. Furthermore, anti-Pol II ChIP-seq experiments demonstrate a redistribution of Pol II molecules into DoG regions accompanied by a decrease in Pol II occupancy at repressed genes upon stress. These results further our understanding regarding the transcriptional landscapes that accompany DoG induction in human cells exposed to hyperosmotic stress.

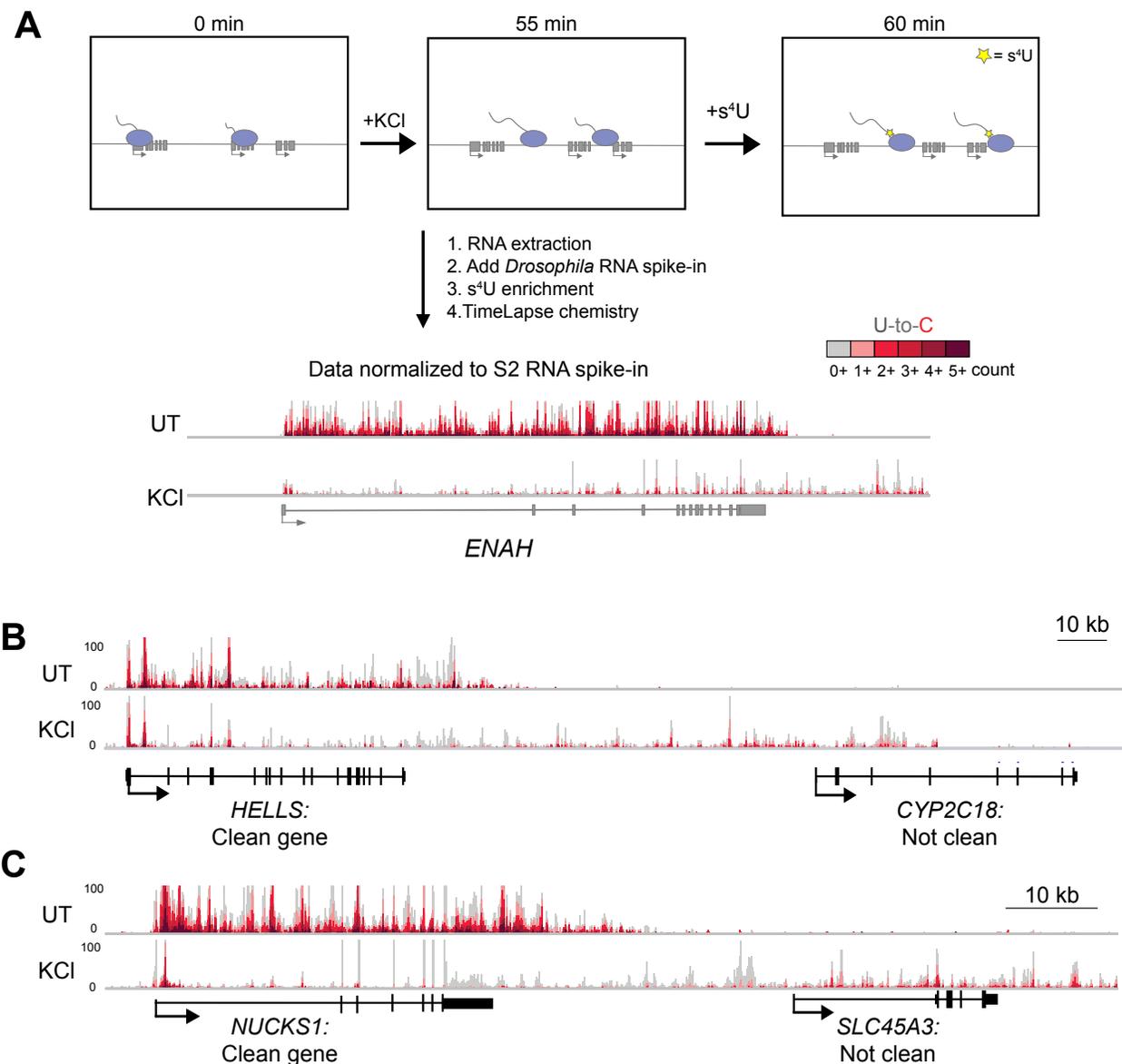


Figure 2.1: TT-TL-seq reveals transcriptional profiles that accompany DoG induction after hyperosmotic stress. A) Setup for TT-TimeLapse sequencing (TT-TL-seq) experiments in HEK-293T cells. An arrow indicating directionality marks the beginning of each transcription unit. Exons are shown as rectangles and Pol II molecules are light purple ovals with attached nascent RNAs. Genome browser views of TT-TL-seq data for *ENAH* provide an example of results from untreated (UT) and KCl-treated (KCl) cells after normalization to the spike-in control. B) Browser image of TT-TL-seq data exemplifying a clean gene (*HELLS*) and a gene that does not meet the criteria for a clean gene (*CYP2C18*). The DoG produced from *HELLS* reads into *CYP2C18*, making the latter appear to be transcriptionally activated by hyperosmotic stress (\log_2 FC=5.39). C) The DoG produced from *NUCKS1* is assigned to *SLC45A3* because of extensive read-in transcription, which also complicates accurate differential expression analysis for *SLC45A3* (\log_2 FC=4.13). ~4-5 kb upstream of *HELLS* and *NUCKS1* are shown.

Results

Hyperosmotic stress causes widespread transcriptional repression

In collaboration with Joshua Zimmer in Professor Matthew Simon's lab, I established the nascent transcriptional profiles accompanying DoG induction by performing TT-TL-seq (Schofield et al. 2018) of untreated HEK-293T cells and cells exposed to hyperosmotic stress (Figure 2.1A). Specifically, I exposed cells to 80 mM KCl for 60 minutes, but added the nucleoside analog 4-thiouridine (s⁴U) during the last 5 minutes to label RNAs being actively transcribed (Schwalb et al. 2016). After extracting RNA from HEK-293T cells, RNA from *Drosophila* S2 cells was added to each sample as a normalization control to ensure accurate differential expression analysis (Lovén et al. 2012; Chen et al. 2015). RNAs containing s⁴U were then biotinylated using methanethiosulfonate chemistry and enriched on streptavidin beads (Duffy et al. 2015). Finally, apparent U-to-C mutations were induced using TL chemistry to assess the nascent nature of the enriched RNAs (Schofield et al. 2018).

It was previously observed that read-in transcription of DoGs into neighboring genes leads to the mis-characterization of overlapping transcripts as being activated by stress (Rutkowski et al. 2015; Hennig et al. 2018; Cardiello et al. 2018; Roth et al. 2020; Figure 2.1B). Moreover, read-in transcription confounds the assignment of DoGs to the corresponding parent gene (Figure 2.1C). Therefore, to ensure accurate differential expression analyses and DoG characterization, I generated a sub-list of genes (referred to as “clean genes”). The term “clean genes” describes genes that are expressed, do not overlap with readthrough regions that correspond to neighboring genes on either strand and have higher expression within the gene body than the region 1 kb upstream of the

gene's transcription start site (TSS) (Roth et al. 2020; Figures 2.1B & C). I identified 4584 clean genes in HEK-293T cells after hyperosmotic stress and analyzed their transcriptional regulation.

Consistent with previous reports analyzing steady-state RNAs in human cells (Amat et al. 2019), changes in transcriptional responses after hyperosmotic stress were observed. TT-TL-seq results reveal predominantly decreases in nascent transcript levels after stress (Figures 2.2A-C). Specifically, the number of normalized read counts corresponding to clean genes decreases 3-fold after KCl treatment (Figure 2.3C). Yet, a subset of clean genes bypasses this transcriptional repression (Figures 2.2B, C & 2.3D), including *GADD45B* (Figure 2.2D), which is known to be induced by hyperosmotic stress (Mak and Kültz 2004). More than 88% of clean genes were repressed after hyperosmotic stress, while only 3% were upregulated (Figures 2.2C & 2.3E). To validate these observations, I extracted total RNA from untreated and KCl-treated cells and performed RT-qPCR using primers targeting intronic regions of several repressed genes. Results from HEK-293T cells and from SK-N-BE(2)C cells confirm the transcriptional decrease of three representative genes upon hyperosmotic stress (Figures 2.3A & B) and demonstrate that mature mRNA levels for these genes remain unaffected, while levels of readthrough transcripts increase after hyperosmotic stress. Together, these results reveal that hyperosmotic stress alters the nascent transcription profiles of cells by repressing thousands of genes and activating a small subset of genes.

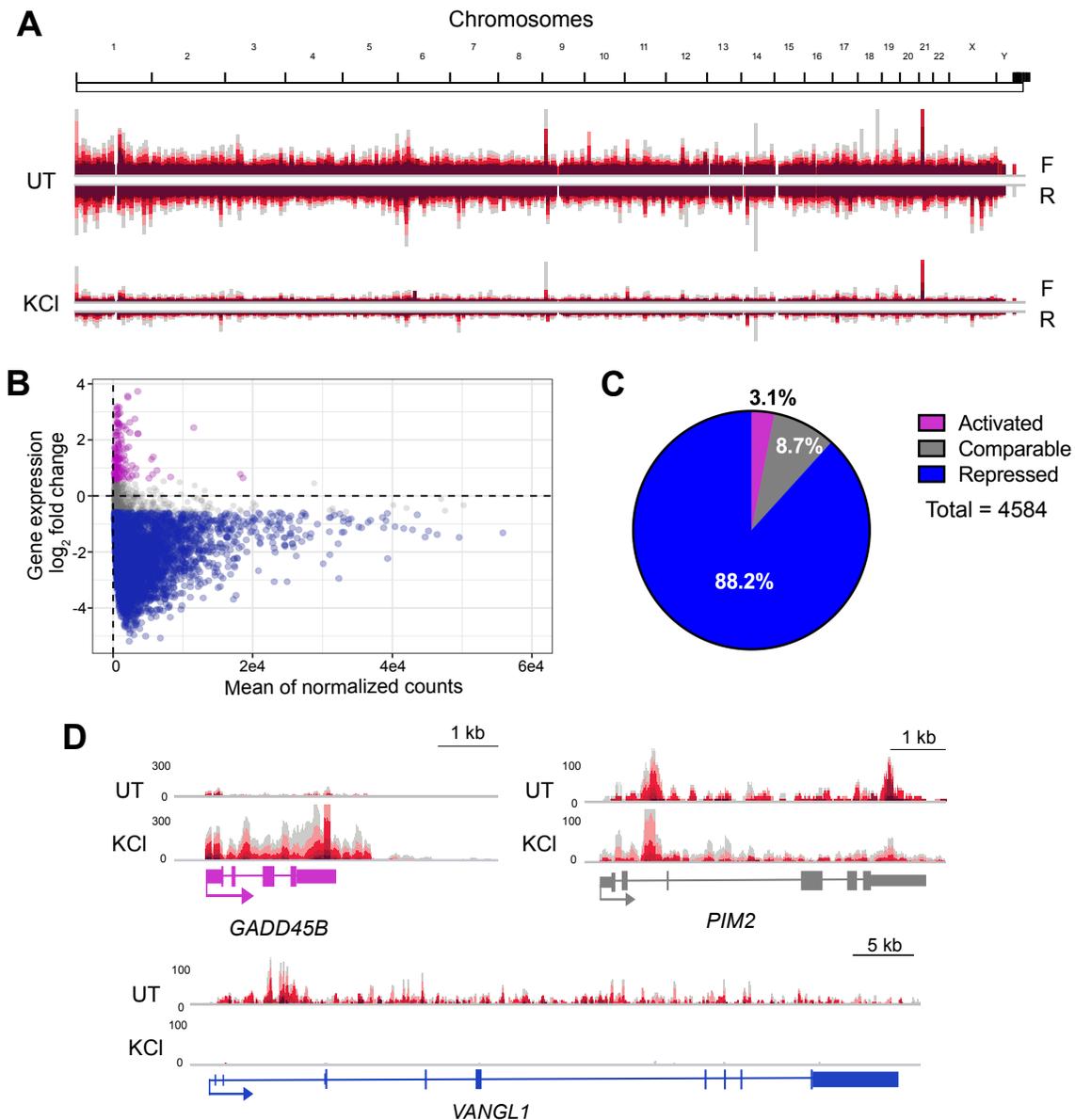


Figure 2.2: Hyperosmotic stress leads to widespread transcriptional repression. A) Whole-genome view of TT-TL-seq normalized reads for forward (F) and reverse (R) strands in UT and KCI samples. B) Minus average plot showing the \log_2 fold change for clean genes on the y-axis and the mean of normalized counts on the x-axis. Activated genes are shown in purple, genes retaining comparable expression are gray and repressed genes are blue (n=4584). C) Pie chart illustrating the percentage of clean genes within each of the 3 categories of transcriptional regulation (activated gene, \log_2 FC > 0.58; comparable gene, \log_2 FC < 0.58 but > -0.58; repressed gene, \log_2 FC < -0.58). D) Browser shots of TT-TL-seq tracks from HEK-293T cells for *VANGL1*, which is transcriptionally repressed by hyperosmotic stress (\log_2 FC = -4.97), *PIM2*, which retains comparable expression after KCI treatment (\log_2 FC = 0.28), and *GADD45B*, which is activated by hyperosmotic stress (\log_2 FC = 3.72).

Stress-induced readthrough transcripts arise independent of gene transcription levels

Consistent with widespread transcriptional repression, normalized TT-TL-seq read counts within the bodies of DoG-producing clean genes decreased after hyperosmotic stress, while read counts corresponding to DoG regions increased (Figure 2.4A). However, \log_2 fold changes (FC) in nascent RNAs of DoG-producing clean genes show that DoGs are produced from genes that experience all three types of transcriptional responses (Figure 2.4B). Specifically, 2.9% of DoGs arise from activated clean genes, 87.8% arise from repressed clean genes and 9.3% arise from clean genes that retain comparable expression in stressed and unstressed HEK-293T cells (Figure 2.4B). I then asked whether DoGs preferentially arise from genes that are transcriptionally repressed upon hyperosmotic stress. Interestingly, the percentage of DoG-producing genes within each class of transcriptional regulation is consistent, comprising 12-14% (Figure 2.4C). Therefore, I conclude that DoGs are produced regardless of the transcriptional level of their upstream genes (Figure 2.4D).

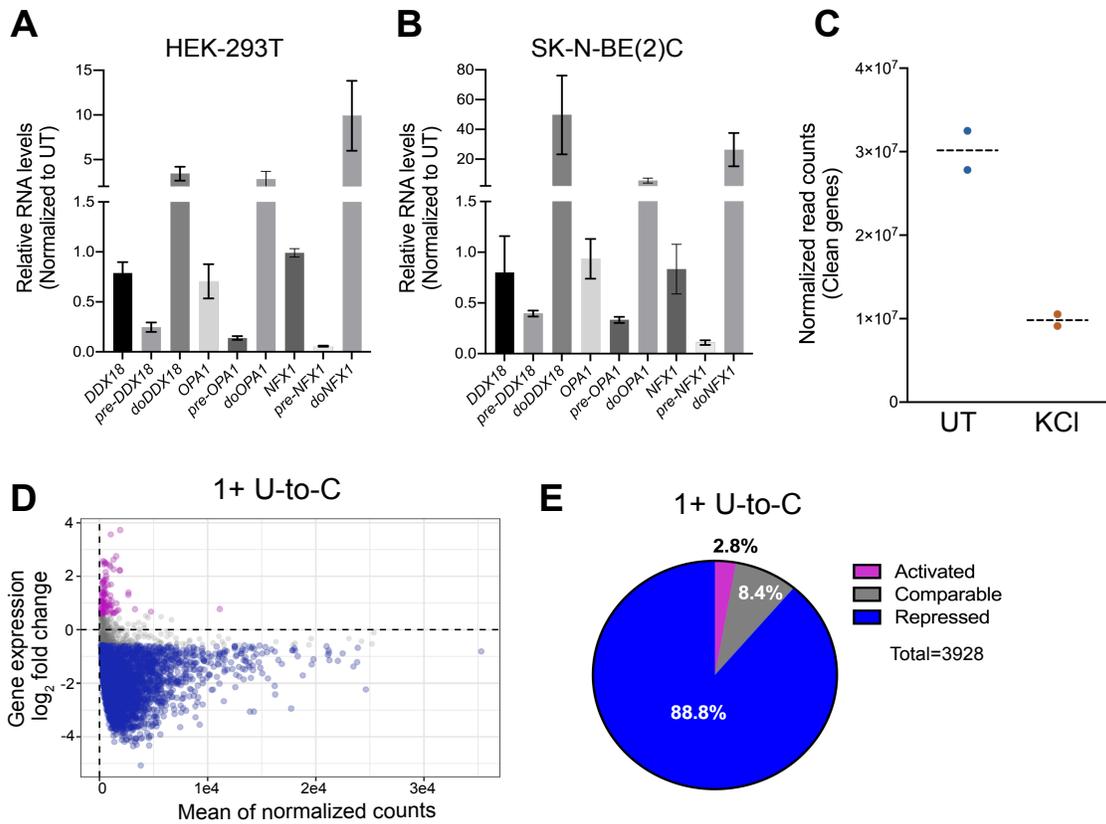


Figure 2.3: Hyperosmotic stress leads to widespread transcriptional repression. Related to figures 1 and 2. A, B) Levels of mRNA, pre-mRNA and DoG RNA for *DDX18*, *OPA1* and *NFX1*, as determined by RT-qPCR in A) HEK-293T cells (n=3) compared to B) SK-N-BE(2)C cells (n=3). C) Interleaved scatter plot showing the sum of all normalized counts for clean genes in UT and KCI samples. D) Minus average plot showing reads with one or more U-to-C mutation induced by TimeLapse chemistry in untreated and KCI-treated samples. E) Pie chart illustrating the percentage of clean genes within each of the 3 categories of transcriptional regulation based on reads containing one or more U-to-C mutation induced by TimeLapse chemistry.

Clean DoG-producing genes are functionally enriched for transcriptional repression

Previous analyses of DoG-producing genes did not reveal any functional enrichment (Vilborg et al. 2017). I suspected that the challenge of assigning DoGs to the correct gene of origin because of their extension into neighboring genes may have complicated previous efforts (Figure 2.1C). Therefore, I revisited the question of whether DoG-producing genes are enriched for certain biological processes using only clean genes. I performed gene ontology analysis of DoG-producing clean genes and clean genes that fail to generate DoGs (non-DoG genes) using Enrichr (Chen et al. 2013; Kuleshov et al. 2016). Interestingly, 5 out of the 10 enriched terms with the most significant p-values for DoG-producing genes are related to transcriptional repression (Figure 2.4E). The remaining 5 terms are related to transcriptional regulation and protein modifications. Non-DoG genes do not show such a striking enrichment for terms related to transcriptional repression compared to other terms (Figure 2.4F). Instead, these genes are strongly enriched for general processes related to RNA processing.

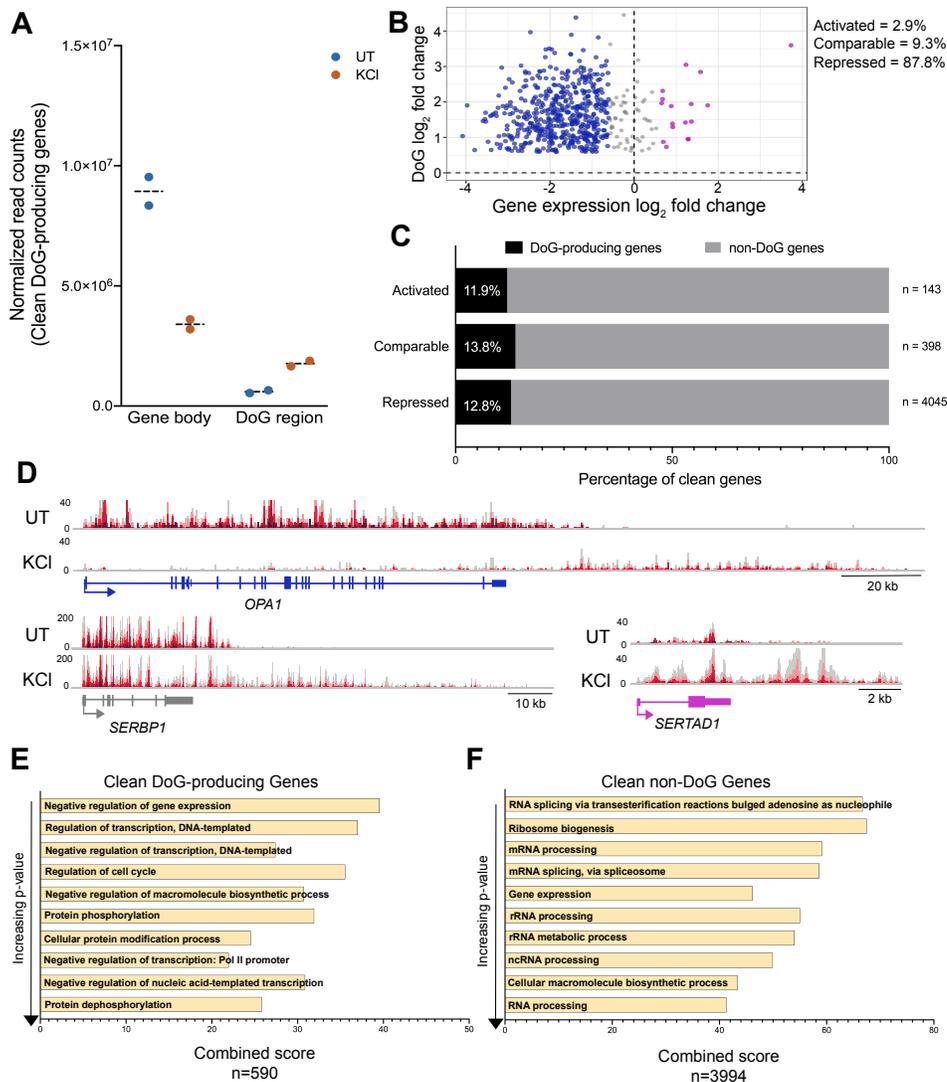
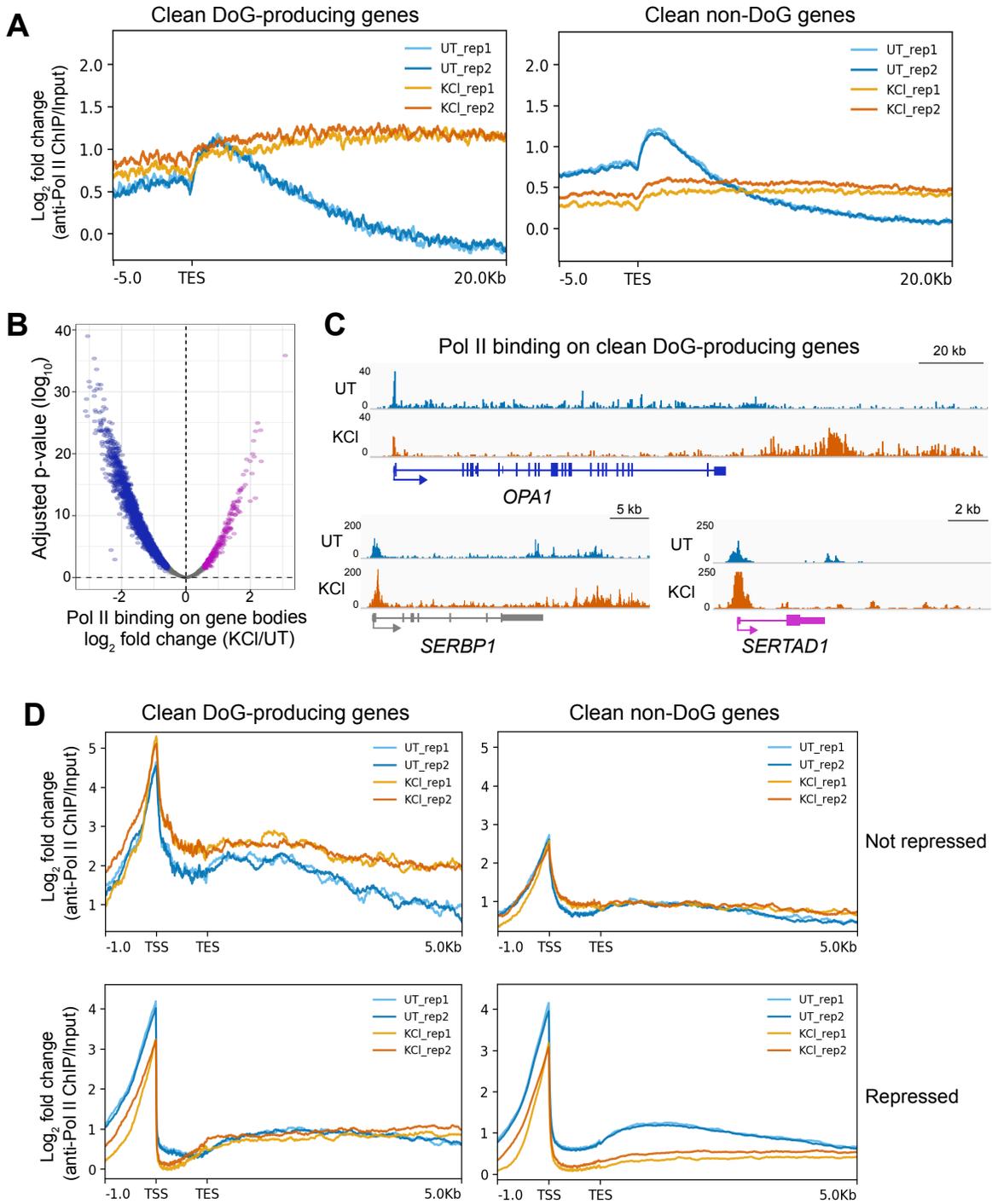


Figure 2.4: DoGs arise regardless of the transcriptional levels of their upstream genes upon hyperosmotic stress. A) Interleaved scatter plot showing the sum of normalized TT-TL-seq read counts of DoG-producing clean genes and corresponding DoG regions ($n=590$) in untreated and KCl-treated HEK-293T cells for two biological replicates. B) Scatter plot showing clean gene \log_2 fold change (FC) for the gene body on the x-axis and the \log_2 FC for the DoG region on the y-axis. DoG-producing genes that are transcriptionally activated upon stress are represented in purple, genes retaining comparable levels of expression are gray and genes that are repressed are blue. C) Bar graph showing the percentage of DoG-producing clean genes (black) within each category of transcriptional regulation. D) Browser image showing UT and KCl TT-TL-seq reads for *OPA1*, a transcriptionally repressed DoG-producing clean gene (gene \log_2 FC= -3.22), for *SERBP1*, which retains comparable expression after stress (gene \log_2 FC= -0.45), and for a transcriptionally activated DoG-producing clean gene, *SERTAD1* (gene \log_2 FC= 1.23). E, F) Bar graphs show gene ontology combined scores for the 10 most significantly enriched biological processes in order of increasing p-value for E) DoG-producing clean genes and for F) non-DoG clean genes.

Pol II is redistributed along the genome after hyperosmotic stress

I used anti-Pol II ChIP-seq to investigate how Pol II binding correlates with the transcriptional profiles observed through TT-TL-seq in untreated and KCl-treated HEK-293T cells. Meta-analyses of clean DoG-producing genes confirmed a redistribution of Pol II molecules to the downstream regions in stressed cells (Figures 2.5A & 2.6A), consistent with DoG transcription and previous observations (Cardiello et al. 2018; Heinz et al. 2018). As expected, this redistribution of Pol II molecules was not observed downstream of non-DoG genes (Figures 2.5A & 2.6A). Interestingly, these results reveal a decrease in Pol II occupancy near the TES of non-DoG genes after hyperosmotic stress that was not observed for DoG-producing genes.

Figure 2.5: Hyperosmotic stress causes redistribution of Pol II molecules across the genome. A) Meta plots showing the \log_2 fold change (FC) of anti-Pol II ChIP-seq data normalized to input across the annotated transcription end sites (TES) of DoG-producing clean genes ($n=590$) and of clean non-DoG genes ($n=3994$) from UT and KCl-treated HEK-293T cells. B) Volcano plot showing the \log_2 FC of read counts from anti-Pol II ChIP-seq data normalized to input for clean genes on the x-axis and the corresponding \log_{10} adjusted p-values on the y-axis. Genes with increased Pol II binding across the gene body are shown in purple, genes retaining comparable levels of Pol II binding are gray and those with decreased Pol II binding across the gene body are blue. C) Browser images of DoG-producing clean genes showing anti-Pol II ChIP-seq tracks normalized to input: *OPA1* is a repressed gene, *SERBP1* retains comparable expression and *SERTAD1* is activated by hyperosmotic stress according to TT-TL-seq data. D) Meta plots showing Pol II binding on DoG-producing clean genes and clean non-DoG genes that are not repressed (top) or genes that are repressed (bottom). Two biological replicates are shown in each meta plot for UT and KCl samples.



I asked to what extent the stress-induced redistribution of Pol II molecules into the regions downstream of genes was responsible for the widespread transcriptional repression observed by TT-TL-seq. Alternatively, the decrease in transcription upon stress might be caused by lowered Pol II binding to repressed genes, as was previously reported for hyperosmotic stress induced by NaCl (Amat et al. 2019). Results show a similar number of Pol II binding peaks after KCl treatment (Figure 2.6B): a subset of overlapping peaks show increased Pol II binding after stress, while a majority of the identified peaks either retain comparable levels or exhibit reduced Pol II occupancy (Figure 2.6C). Consistently, I found that Pol II binding decreases over most clean gene bodies after KCl treatment (Figure 2.5B).

I verified Pol II occupancy patterns for representative genes within each category of transcriptional regulation. *OPA1*, a DoG-producing gene that is transcriptionally repressed, shows decreased binding close to the TSS and a shift of Pol II peaks into the downstream region in stressed cells (Figure 2.5C). In contrast, *SERTAD1*, a DoG-producing gene that is activated by hyperosmotic stress, and *SERBP1*, which retains comparable expression according to TT-TL-seq data, show increased Pol II binding close to each TSS that extends downstream of these genes. Finally, I assessed the distribution of Pol II molecules across repressed clean genes and across clean genes that are either activated or retain comparable expression after hyperosmotic stress (not repressed). Consistent with the individual gene examples discussed above, clean DoG-producing genes that are not repressed by hyperosmotic stress show increased Pol II binding close to their TSSs and across the gene body, while genes that are transcriptionally repressed by hyperosmotic stress show decreased levels of Pol II binding across the gene body.

Similar effects were also observed for clean non-DoG genes (Figure 2.5D). These results are, thus, in agreement with the nascent transcription profiles derived from TT-TL-seq (Figures 2.2 & 2.3), supporting a model where hyperosmotic stress induces both redistribution of Pol II molecules along the genome and a decrease in Pol II binding to repressed genes.

Discussion

TT-TL-seq was used to evaluate nascent transcription profiles that accompany DoG induction under conditions of hyperosmotic stress (Figure 2.1). I observed that hyperosmotic stress triggers transcriptional repression that is widespread and of greater magnitude than anticipated (Robbins et al. 1970; Amat et al. 2019). After an hour-long exposure to mild conditions of hyperosmotic stress, only ~12% of clean genes escaped transcriptional repression (Figures 2.2 & 2.3). Interestingly, DoGs are produced independent of the transcriptional levels of their upstream genes, and DoG-producing genes are functionally enriched for transcriptional repression according to gene ontology analyses (Figure 2.4). Pol II binding profiles demonstrate a correlation with nascent transcription profiles across DoG-producing clean genes and non-DoG genes (Figure 2.5).

Analysis of clean genes (expressed genes lacking overlap with readthrough regions) enabled me to accurately characterize the relationship between DoGs and their upstream transcripts (Figures 2.1 & 2.4). Consistent with previous observations in heat-shocked cells (Cardiello et al. 2018), I report that DoGs are produced from genes that experience different transcriptional responses to hyperosmotic stress. Importantly, a similar percentage of clean genes that are activated, repressed or unchanged produce DoGs

after hyperosmotic stress, revealing that transcriptional levels and DoG production are not mechanistically linked. My analyses further suggested that many genes expressed in HEK-293T cells might be affected by read-in transcription upon exposure to hyperosmotic stress. Intriguingly, other groups have previously reported that read-in transcription from upstream genes negatively affects the splicing of downstream genes upon HSV-1 infection (Rutkowski et al. 2015). Additionally, there is evidence that heat shock and hyperosmotic stress disrupt splicing (Shalgi et al. 2014; Hennig et al. 2018). These findings suggest that altered splicing patterns may be a consequence of DoG production and highlight the extensive impact of DoG production on the transcriptome of stressed cells. Investigating the diversity of unprocessed transcripts that accumulate upon stress will provide a better understanding of how transcriptional and processing dynamics are impacted by stress.

Previously, our lab had been unable to find any functional enrichment for DoG-producing genes in the context of hyperosmotic stress (Vilborg et al. 2017). Upon revisiting this question using only clean genes, I found that many clean DoG-producing genes encode transcriptional repressors (Figure 2.4). Transcriptional repression has also been reported under other DoG-inducing conditions, including heat shock and viral infection (Mahat et al. 2016; Rutkowski et al. 2015; Bauer et al. 2018). Thus, it is tempting to speculate that DoGs could serve as a store of unprocessed transcripts to facilitate cell survival upon prolonged stress by providing a source of important mRNAs (requiring processing, but not active transcription), thereby reducing the cell's energetic costs. Other terms of biological enrichment for DoG-producing genes are related to protein modifications. A repository of transcripts that encode such proteins might facilitate

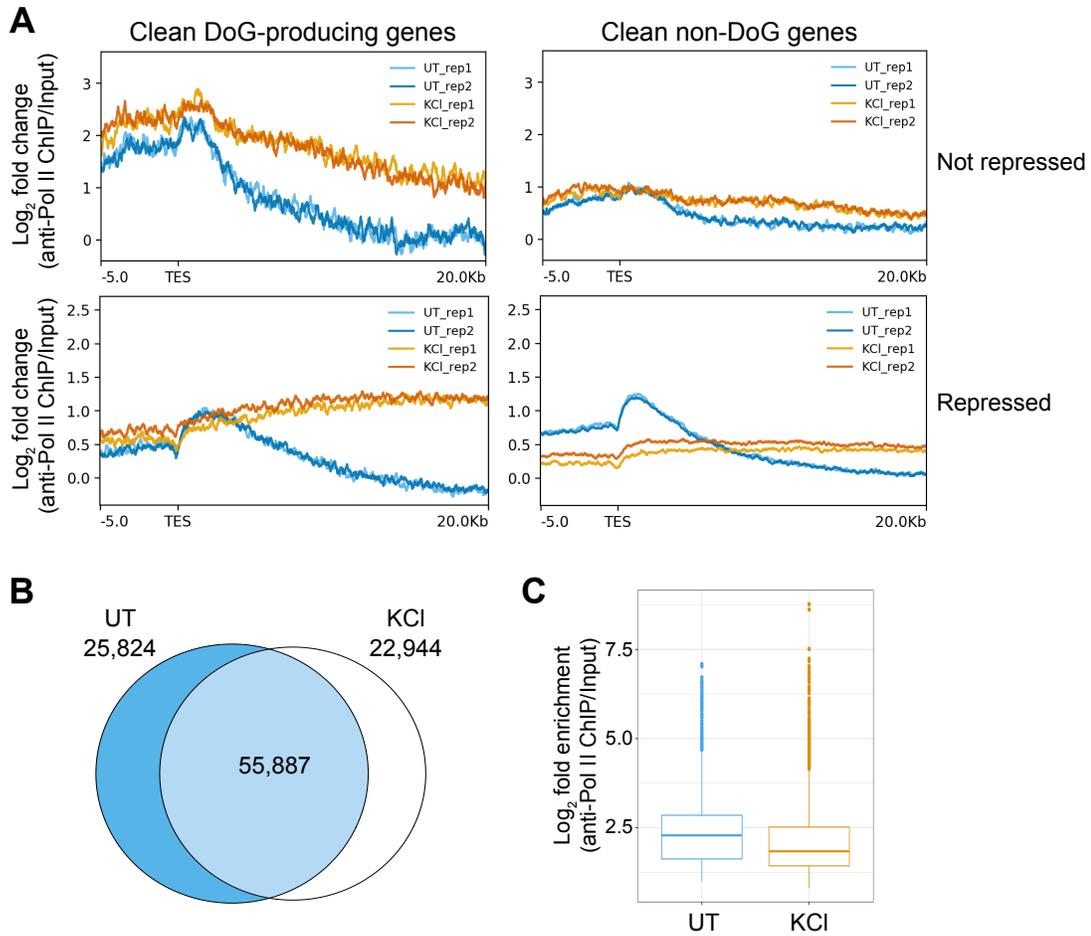


Figure 2.6: Pol II is redistributed along the genome upon hyperosmotic stress. Related to figure 4. A) Meta-plots showing the log₂ fold change of anti-Pol II ChIP-seq data normalized to input across the annotated transcription end sites (TES) for DoG-producing clean genes and clean non-DoG genes that are not repressed (top) and for clean genes that are repressed upon hyperosmotic stress (bottom). Two biological replicates from untreated and KCl-treated HEK-293T cells are shown in each plot. B) Venn diagram showing Pol II binding peaks detected in UT and KCl samples (55,887). 22,944 peaks were detected exclusively in the KCl samples, while 25,824 peaks were found only in the UT samples. C) Box plot depicting the log₂ fold enrichment (ChIP/Input) of Pol II peaks in UT and KCl samples.

recovery from stress. Nuclear-retained, intron-containing transcripts have been shown to serve as a source of gene regulation in other contexts (Boutz et al. 2015). Therefore, there is precedent for unprocessed transcripts, such as DoGs, functioning to regulate gene expression levels via post-transcriptional processing. This idea is also supported by the fact that DoGs have relatively long half-lives (Vilborg et al. 2015). DoG induction might further promote transcriptional repression during hyperosmotic stress by causing transcriptional interference at neighboring genes (Mazo et al. 2007; Muniz et al. 2017). Alternatively, given the transcriptional changes observed upon stress, DoGs may serve to alter the chromatin landscape surrounding their parent and neighboring genes to facilitate prolonged responses throughout stress and recovery from stress.

In agreement with previous observations in heat shock (Cardiello et al. 2018), anti-Pol II ChIP-seq data from KCl-treated cells demonstrate a decrease in Pol II binding across the bodies of repressed clean genes accompanied by a shift of elongating Pol II molecules into downstream regions (Figure 2.5). At loci that are not repressed by stress, I observe an increase in Pol II binding along the gene bodies, extending into their downstream regions. These observations are consistent with the TT-TL-seq results discussed above (Figures 2.2 & 2.3). The differing levels of Pol II occupancy observed across parent genes suggest an elegant and tightly regulated mechanism of transcriptional termination at DoG-producing genes. Furthermore, my anti-Pol II ChIP-seq data revealed decreased occupancy near the TES of non-DoG genes after hyperosmotic stress, which was not observed for DoG-producing genes (Figs. 2.5A & 2.6A). This observation suggests that one factor distinguishing DoG-producing genes from genes that do not produce DoGs is their dependence on Pol II pausing near their TESs for efficient transcription termination.

Chapter 3

Hyperosmotic stress alters the RNA Polymerase II interactome

Summary

The mechanisms underlying DoG formation upon exposure of human cells to environmental stressors remain unknown. In collaboration with Dr. Maria Apostolidi in Professor Jesse Rinehart's lab, I performed mass spectrometry and western blot experiments on untreated and KCl-treated cells to investigate alterations to the RNA Pol II interactome. Results revealed that while certain cleavage and polyadenylation factors remain Pol II-associated, Integrator complex subunits dissociate from Pol II under stress, leading to a genome-wide loss of Integrator on chromatin. Depleting the catalytic subunit of the Integrator using siRNAs induces hundreds of readthrough transcripts, whose parental genes partially overlap those of stress-induced DoGs. These results provide insights into the molecular mechanism(s) of DoG production and reveal an unanticipated role for the Integrator complex in the production of these long noncoding RNAs.

The work presented in this chapter has been published in *Molecular Cell* (Rosa-Mercado et al. 2021).

Introduction

The production of DoGs is a consequence of decreased transcription termination (Vilborg et al. 2015). Upon influenza and HSV-1 infection, DoG induction is orchestrated in part by viral proteins NS1 and ICP27, respectively. These proteins interfere with transcription termination by interacting with host cleavage and polyadenylation (CPA) factors (Nemeroff et al. 1998; Bauer et al. 2018; Wang et al. 2020; Figure 1.3). Accordingly, knockdown of the catalytic subunit of the CPA complex, CPSF73, partially induces DoGs (Vilborg et al. 2015). However, host-dependent mechanisms underlying DoG production upon environmental stress, such as hyperosmotic stress, remain unknown.

The Integrator complex is important for the 3'-end processing of various noncoding RNAs, such as small nuclear RNAs (snRNAs), enhancer RNAs, viral microRNAs, and lncRNAs, as well as replication-dependent histone pre-mRNAs (Albrecht and Wagner 2012; Lai et al. 2015; Xie et al. 2015; Nojima et al. 2018; Skaar et al. 2015). Integrator also regulates promoter-proximal Pol II pausing at protein-coding genes (Stadelmayer et al. 2014; Gardini et al. 2014; Elrod et al. 2019; Beckedorff et al. 2020). Integrator 11 (Int11) is the catalytic subunit of the complex and forms a heterodimer with Int9 (Baillat et al. 2005; Wu et al. 2017). These subunits are analogous to CPSF73 and CPSF100 (Dominski et al. 2005; Baillat and Wagner 2015), respectively, which are essential for the 3' processing of mRNAs (Mandel et al. 2006; Shi et al. 2009).

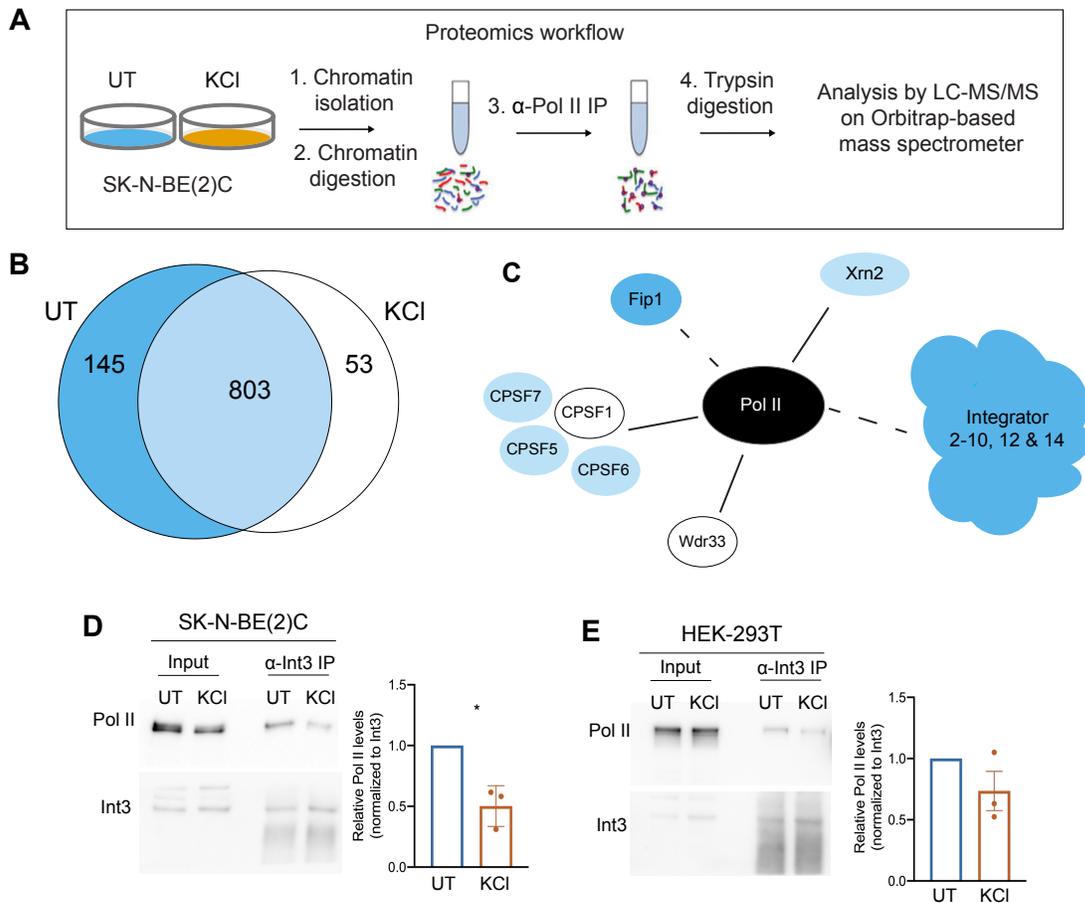


Figure 3.1: The Pol II interactome changes upon hyperosmotic stress. A) Experimental setup for mass spectrometry of anti-Pol II immunoprecipitates obtained from the chromatin fractions of SK-N-BE(2)C cells. B) Overlap between Pol II-associated proteins found in UT and KCl-treated samples (803). Peptides for 145 proteins were found only in the UT samples (blue) and peptides corresponding to 53 proteins were found only in the KCl-treated samples at least once across six replicates (white). C) Cartoon depiction of changes in Pol II interactions with termination factors. Solid lines represent interactions that remain after hyperosmotic stress, while dashed lines represent interactions found only in the untreated samples. Proteins detected only in untreated samples are shown in blue, proteins present only in the KCl-treated samples are white and proteins present in both samples are light blue. D, E) Western blots of anti-Int3 immunoprecipitates obtained from chromatin fractions reveal decreased levels of associated Pol II in KCl-treated samples compared to untreated samples in C) SK-N-BE(2)C (p-value=0.0357) and in D) HEK-293T cells (p-value=0.2407). Chromatin fractions from these cell lines are shown as input. Around 8% of the total input and 11% of the immunoprecipitates were loaded on the gels.

Mass spectrometry experiments performed in collaboration with Dr. Maria Apostolidi in Professor Jesse Rinehart's lab reveal changes in the Pol II interactome after stress. Specifically, we observe a decreased interaction between Pol II and subunits of the Integrator complex, while the interactions between Pol II and factors known to mediate the termination of protein-coding genes are unchanged. ChIP-seq experiments reveal a genome-wide decrease of Int11 and Int3 occupancy on DNA. Correspondingly, siRNA knockdown of Int11 was sufficient to induce readthrough transcription at hundreds of genes. The identity of genes that produce readthrough transcripts upon depletion of functional Int11 partially overlaps the collection of genes that give rise to DoGs after hyperosmotic stress. Together, these results provide mechanistic insights into the biogenesis of a recently characterized class of noncoding RNAs and reveal a novel role for the Integrator complex in DoG biogenesis.

Results

The Pol II interactome is altered by hyperosmotic stress

Mass spectrometry analyses of anti-Pol II immunoprecipitates obtained from the chromatin fractions of untreated or KCl-treated SK-N-BE(2)C cells were performed in order to gain insight into the mechanisms that promote DoG formation upon hyperosmotic stress (Harlen et al. 2016; Figure 3.1A). This human neuroblastoma cell line, where KCl-induced DoGs were originally described (Vilborg et al. 2015), was used because I observed more robust DoG induction in these cells than in HEK-293T cells (Figures 2.3A & B). I confirmed isolation of chromatin through western blots (Figure 3.2A). Overall, mass spectrometry results reveal distinct changes in the Pol II interactome (Figure 3.1B). Peptides corresponding to several CPA factors, as well as to Xrn2 were detected in both

untreated and KCl-treated samples at least once across six experiments (Figure 3.1C). The presence of Xrn2 and CPSF1 in the anti-Pol II immunoprecipitates of untreated and KCl-treated cells was validated through western blots (Figures 3.2B-D). In contrast, 11 of the 14 subunits of the Integrator complex were not detected among the Pol II interactors in the KCl-treated samples (Figure 3.1C). To validate this observation, I performed western blots to detect Pol II in chromatin-bound anti-Integrator subunit 3 (Int3) immunoprecipitates from SK-N-BE(2)C and HEK-293T cells. Decreased binding between Pol II and Int3 in KCl-treated samples compared to untreated samples was confirmed for both cell lines (Figures 3.1D & E). The specificity of the anti-Int3 antibody was verified by performing siRNA-mediated knockdowns of this subunit followed by western blots of HEK-293T whole cell lysates (Figure 3.2E). To ensure that the decreased detection of Integrator subunits was not due to protein degradation, I also assessed the levels of Int3 and Int11 in whole cell lysates from SK-N-BE(2)C cells. Western blots confirmed that the levels of these Integrator subunits do not decrease in cells after hyperosmotic stress (Figure 3.2F).

Together, proteomic results show that hyperosmotic stress induces changes to the Pol II interactome that could contribute to the transcriptional landscape revealed by high throughput sequencing. Specifically, important termination factors interact with Pol II despite cellular stress, while the interaction between Pol II and the Integrator complex decreases.

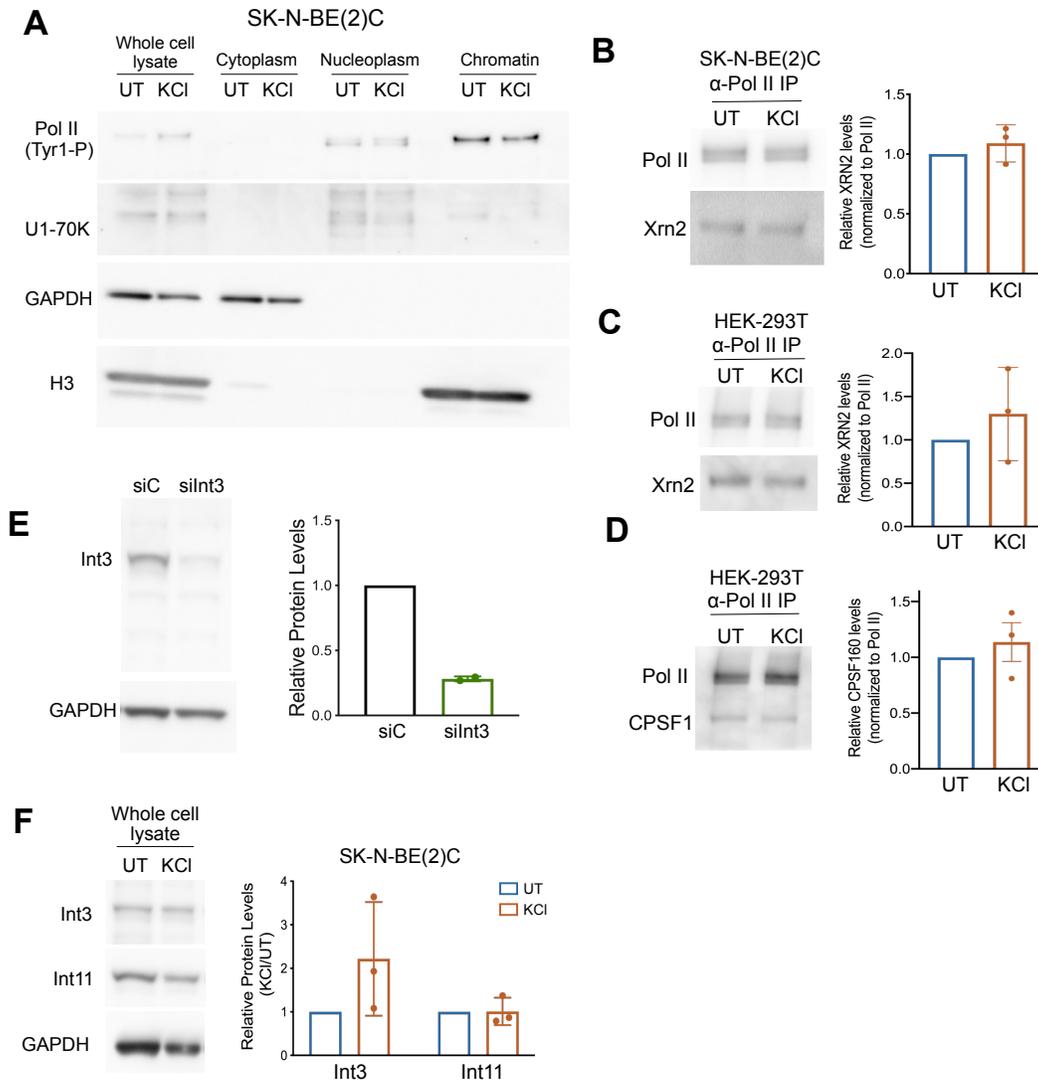


Figure 3.2: Levels of certain termination factors remain unchanged after hyperosmotic stress. Related to figure 5. A) Western blots of cellular fractions from SK-N-BE(2)C cells. Histone H3 and Pol II phosphorylated on Tyr1 residues of the Pol II CTD served as chromatin markers, U1-70K as a nucleoplasmic marker and GAPDH as a cytoplasmic marker. B, C) Western blots of anti-Pol II immunoprecipitates from B) SK-N-BE(2)C cells (p-value = 0.4217) and C) HEK-293T cells confirm that Pol II binding to Xrn2 does not decrease after hyperosmotic stress (p-value = 0.4383). D) Western blotting of anti-Pol II immunoprecipitates using an antibody against CPSF1 confirms that hyperosmotic stress does not alter the binding between CPSF1 and Pol II (p-value = 0.5128). E) Western blot of whole cell lysates demonstrating a decrease in Int3 detection by the anti-Int3 antibody used throughout this study in HEK-293T cells transfected with an siRNA targeting this subunit of the Integrator complex (siInt3) compared to cells transfected with a scrambled siRNA control (siC). Band intensities for Int3 normalized to GAPDH were quantified (n=2). F) Western blot of SK-N-BE(2)C whole cell lysates detecting comparable levels of Int3 and Int11 in UT and KCI-treated samples.

Hyperosmotic stress decreases the occupancy of Integrator subunits on DNA

Further insight into the decreased binding of Integrator subunits to Pol II after hyperosmotic stress (Figure 3.1) was gained by performing anti-Int11 and anti-Int3 ChIP-seq in untreated and KCl-treated HEK-293T cells. The Integrator complex was previously shown to bind close to the TSS of genes, where it regulates promoter proximal Pol II pause release (Stadelmayer et al. 2014; Elrod et al. 2019; Beckedorff et al. 2020). Consistently, Int11 and Int3 peaks were predominantly observed close to the TSS of clean genes that decrease after hyperosmotic stress (Figures 3.3A & B). Meta-analysis of DoG-producing clean genes and clean non-DoG genes revealed similar decreases in Int11 and Int3 occupancy close to the TSSs. Additionally, read counts for identified Int11 binding sites confirm decreased occupancy of this protein on DNA after hyperosmotic stress (Figure 3.3C).

Since TT-TL-seq and anti-Pol II ChIP-seq experiments reveal that hyperosmotic stress causes transcriptional repression accompanied by a decrease in Pol II binding at downregulated genes, I asked whether the observed decreased occupancy of Integrator subunits on DNA was a consequence of a decrease in Pol II occupancy. The occupancy of Int3 and Int11 decreased near the TSS of genes that are not repressed by hyperosmotic stress, including *SERTAD1* and *SERBP1* (Figures 2.3D, 2.4C & 3.4A). Furthermore, I quantified the occupancy of Int11 and Pol II across Int11 binding sites that were identified across two biological replicates and observed that the decrease in Int11 binding is not limited to sites exhibiting a concomitant decrease in Pol II occupancy (Figures 3.4A & B). I conclude that the decreased binding of Integrator subunits to DNA is not a mere consequence of the transcriptional repression induced by hyperosmotic

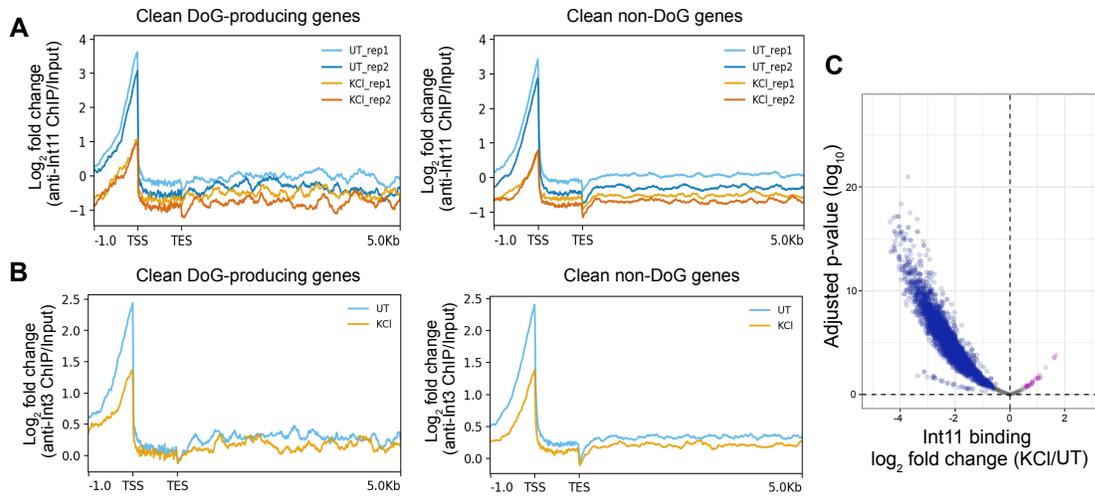


Figure 3.3: Hyperosmotic stress leads to decreased occupancy of Integrator subunits on DNA. A) Meta-plots showing the \log_2 FC of anti-Int11(top) or anti-Int3 ChIP-seq normalized to input reveal decreased occupancy of Int11 near the TSSs of clean DoG-producing genes ($n=590$) and clean non-DoG genes ($n=3994$). B) Volcano plot of read counts for Int11 binding sites. \log_2 FC of KCI/UT is shown on the x-axis with the corresponding adjusted p -values on the y-axis ($n=3881$). Sites where Int11 occupancy decreases after hyperosmotic stress are blue, sites where occupancy remains comparable are gray and sites where occupancy increases are purple.

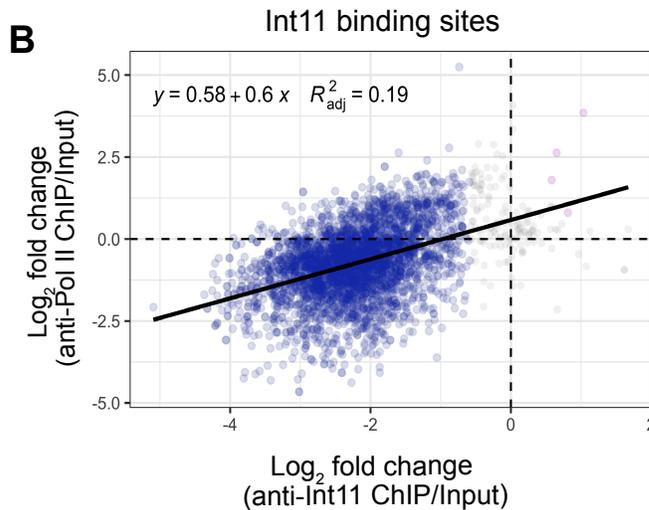
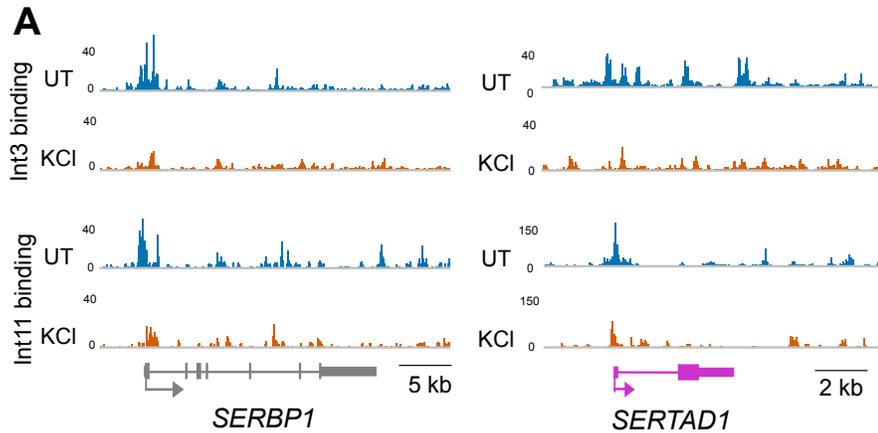


Figure 3.4: Decreased occupancy of Integrator subunits on DNA upon hyperosmotic stress is not solely dependent on a decrease in Pol II binding. A) Browser images of anti-Int3 and anti-Int11 ChIP-seq tracks normalized to input for *SERBP1* and *SERTAD1*, which are DoG-producing clean genes that are not repressed by hyperosmotic stress. B) Scatter plot showing the log₂ fold change of read counts for anti-Pol II ChIP-seq data normalized to input for identified Int11 binding sites on the y-axis and the log₂ fold change of anti-Int11 ChIP-seq data normalized to input on the x-axis (n = 3783). Sites where Int11 occupancy decreases after stress are blue, sites that retain comparable levels of Int11 occupancy are gray and sites with increased Int11 occupancy are purple.

stress.

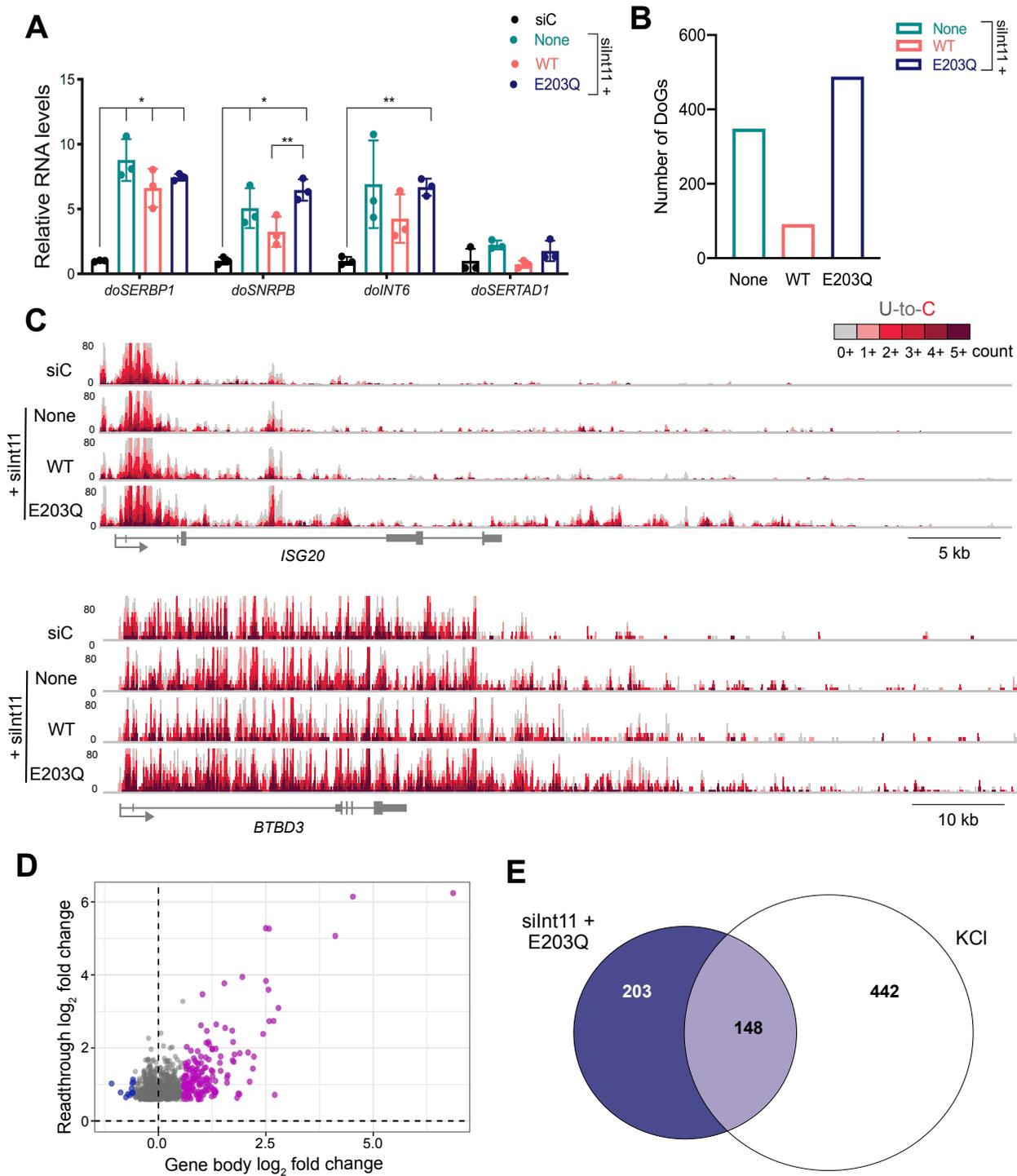
Depletion of Integrator endonuclease leads to DoG production

The Integrator complex regulates transcription termination at many noncoding RNA loci and has been shown to bind the 3'-end of certain protein-coding genes (Baillat and Wagner 2015; Gardini et al. 2014). Since the interaction between Pol II and Integrator decreases after hyperosmotic stress (Figure 3.1), I investigated whether knocking down the catalytic subunit of the complex, Int11, using siRNAs is sufficient to induce DoGs. I transfected HEK-293T cells with an siRNA against Int11 (silInt11) or with a non-targeting siRNA control (siC). HEK-293T cells stably expressing FLAG-tagged, siRNA-resistant wild-type (WT) or catalytically inactive Int11 (E203Q) were also transfected with silInt11 (Baillat et al. 2005; Xie et al. 2015) for 72 hours. DoG levels were then measured by RT-qPCR, confirming that knockdown of endogenous Int11 is sufficient to induce DoGs. Moreover, expression of exogenous WT Int11 reduces DoG induction compared to samples expressing the E203Q mutant or to samples expressing no rescue (Figure 3.5A).

Endogenous Int11 protein levels were reduced by ~90% in silInt11-transfected samples (Figure 3.6A). The efficiencies of exogenous WT and E203Q mutant Int11 expression were verified by measuring the levels of unprocessed U2 through RT-qPCR (Skaar et al. 2015; Figure 3.6B): unprocessed U2 accumulates in cells without rescue (None) or expressing the E203Q mutant Int11, but not in cells transfected with siC or cells expressing exogenous WT Int11. Northern blots using probes against snRNA U2 showed that levels of mature U2 do not decrease after knockdown (Figure 3.6C), consistent with previous reports (Tatomer et al. 2019). Thus, it is unlikely that the effects of Int11 knockdown on DoG production are because of the role of Integrator in snRNA processing.

In collaboration with Joshua Zimmer in Professor Matthew Simon's lab, I assessed the extent to which knockdown of endogenous Int11 induces DoGs genome-wide. To increase cell viability, we performed TT-TL-seq on HEK-293T cells transfected with an siRNA against Int11 for 48 hours (Figure 3.6D). Results obtained from cells lacking functional Int11 reveal hundreds of readthrough sites across the genome that are induced by more than 1.5-fold compared to the siC-transfected sample (Figures 3.5B & D). According to TT-TL-seq data, induction of readthrough transcription after depletion of endogenous Int11 was most evident upon expression of the E203Q mutant (Figures 3.5B & C). Yet, readthrough transcripts observed in silInt11-transfected cells expressing no rescue and in cells expressing the E203Q mutant Int11 were highly correlated (Figure 3.6E). As expected, the most highly induced sites of readthrough transcription corresponded to snRNA genes (Baillat et al. 2005; Albrecht and Wagner 2012). I also detected readthrough downstream of lncRNA and histone genes as previously described (Nojima et al. 2018; Skaar et al. 2015). However, most identified readthrough sites were

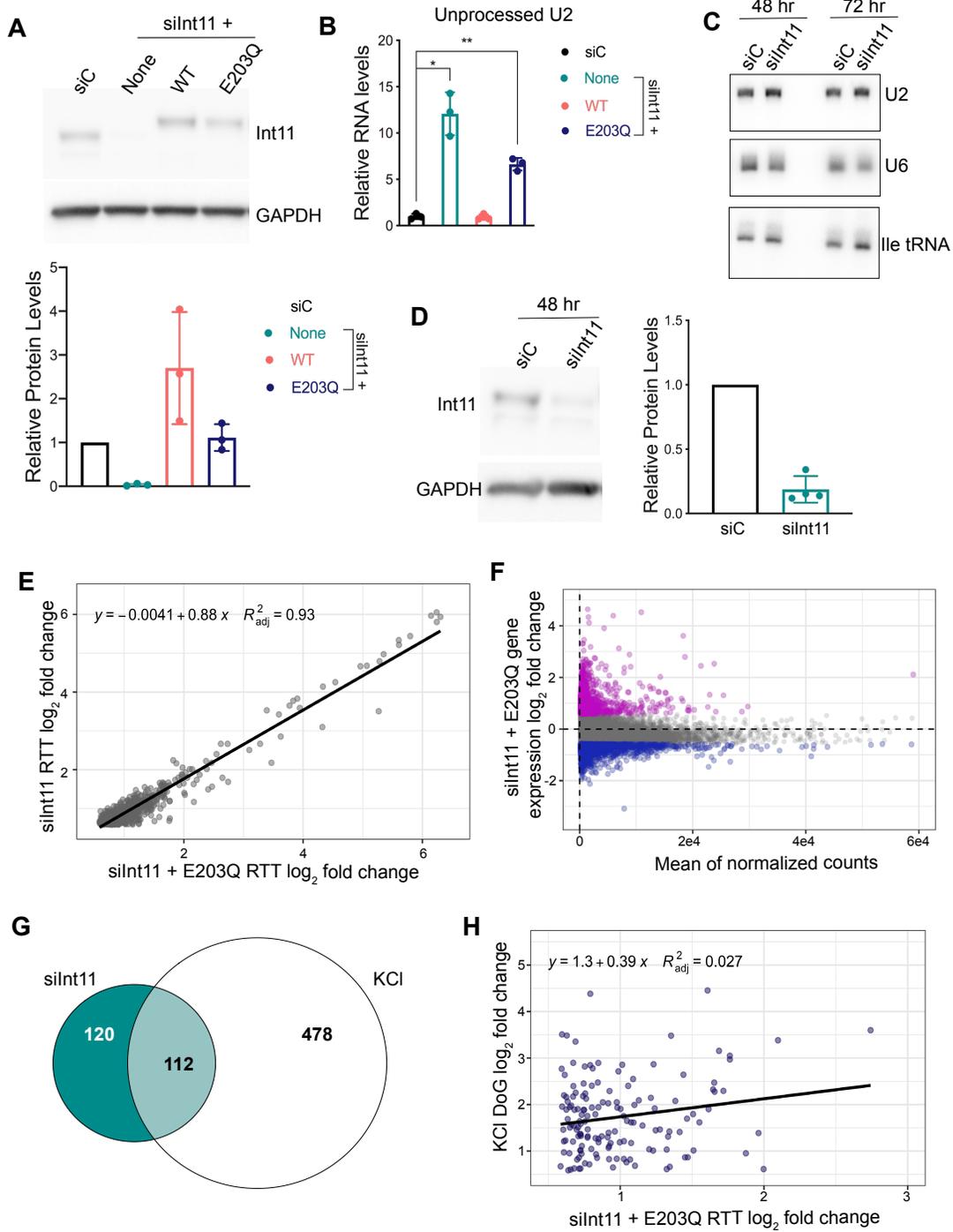
Figure 3.5: Depletion of Integrator nuclease subunit leads to readthrough transcription. A) RT-qPCR analysis of DoG levels in HEK-293T cells transfected with a scrambled siRNA control (siC) or an siRNA targeting endogenous Int11 (silInt11). The silInt11-transfected cells expressed either exogenous wild-type Int11 (WT), catalytically inactive Int11 (E203Q) or no exogenous Int11 (None). Adjusted p-values compared to siC: None: doSERBP1=0.004; doSNRBP=0.032; doInt6=0.077; doSERTAD1=0.097; WT: doSERBP1=0.011; doSNRBP=0.091; doINT6=0.091; doSERTAD1=0.654; E203Q: doSERBP1=<0.001; doSNRBP=<0.001; doINT6=<0.001; doSERTAD1=0.337. B) Bar graph showing the number of DoGs induced after siRNA knockdown of endogenous Int11. C) Browser image of TT-TL-seq data for two genes that produce DoGs upon depletion of functional Int11. D) Scatter plot showing gene expression log₂ fold change (FC) for silInt11+E203Q HEK-293T cells on the x-axis and the log₂ FC of the corresponding readthrough transcripts on the y-axis. Genes that are activated in silInt11+E203Q cells compared to siC-transfected cells are purple, unaffected genes are gray and repressed genes are blue. All readthrough sites identified in the silInt11+E203Q sample are represented in this plot (n=840). E) Venn diagram displaying overlap between the identities of clean genes that produce readthrough transcripts in silInt11+E203Q cells (dark blue) and those of DoG-producing clean genes in KCl-treated samples (white).



downstream of protein-coding genes. Of the 840 readthrough transcripts induced by knockdown of functional Int11 (E203Q sample), 489 exhibited greater than 80% read coverage in the region 5 kb downstream of the annotated termination site of the gene of origin and, therefore, met all criteria to be classified as DoG RNAs (Figures 3.5B & C).

Depletion of Integrator subunits has been shown to alter the transcriptional levels of certain genes (Gardini et al. 2014; Elrod et al. 2019; Tatomer et al. 2019; Beckedorff et al. 2020). Consistently, TT-TL-seq results show that depletion of functional Int11 in HEK-293T cells differentially affects more than a thousand genes (Figure 3.6F). Examination of the expression levels of the parent genes revealed that readthrough transcripts

Figure 3.6: Int11 knockdown leads to readthrough transcription. Related to figure 7. A) Western blot showing the levels of Int11 and GAPDH in lysates from HEK-293T cells transfected with a scrambled siRNA control (siC) or with an siRNA against Int11 (silInt11) and in rescue cell lines stably expressing WT or E203Q Int11 transfected with silInt11 for 72 hours. Bar graph shows the quantification of Int11 band intensities normalized to GAPDH band intensities. B) RT-qPCR results confirm accumulation of unprocessed U2 in silInt11-transfected cells with no rescue (None) (p-value = 0.0124) or with E203Q rescue (p-value = 0.0061), but not in cells with wild-type rescue (WT) (p-value = 0.3431). C) Probes against U2 snRNA, U6 snRNA and Isoleucine tRNA were used for northern blotting of total RNAs extracted from siC and silInt11-transfected samples 48 hours and 72 hours after transfection. D) Western blot showing the levels of Int11 and GAPDH in lysates from HEK-293T cells transfected with a scrambled siRNA control (siC) or silInt11 for 48 hours (81% knockdown efficiency). E) Scatter plot showing correlation between readthrough transcripts detected in silInt11-transfected cells expressing no rescue and cells expressing the E203Q mutant Int11 (n=576). F) Mean average plot showing read counts for expressed genes on the x-axis and their log₂ FC after depletion of functional Int11 (E203Q sample) on the y-axis (n=14,640). Activated genes are shown in purple and repressed genes in blue. G) Venn diagram showing the overlap between the identities of clean genes that produce readthrough transcripts in silInt11-transfected cells (teal) and DoG-producing genes in KCl-treated samples (white). H) Scatter plot demonstrating a correlation between the log₂ FC of overlapping genes in KCl-treated cells (y-axis) and silInt11+E203Q cells (x-axis) (n=148).



predominantly arise from upregulated genes or from genes that retain comparable expression after knockdown, while very few arise from genes that are transcriptionally repressed (Figure 3.5D).

Given the observation that the interaction between Integrator subunits and Pol II is disrupted by hyperosmotic stress, I asked how the identities of genes producing readthrough transcripts upon depletion of functional Int11 compare to genes that produce DoGs after hyperosmotic stress. I identified 232 clean genes producing readthrough transcripts in silInt11-transfected cells and 351 clean genes producing readthrough transcripts in the silInt11+E203Q mutant sample. Comparison with the 590 DoG-producing clean genes identified in stressed cells showed that up to 25% of KCl-induced DoGs are detected at loci that also produce readthrough transcripts after depletion of functional Int11 (Figures 3.5E & 3.6G). These readthrough transcripts are generally more robustly induced after KCl treatment than upon depletion of functional Int11 (Figure 3.6H), suggesting that, although Int11 knockdown is sufficient to produce readthrough transcription, decreased interactions between Integrator and Pol II are not solely responsible for DoG induction upon hyperosmotic stress.

Discussion

I set out to investigate the mechanisms that account for DoG biogenesis by performing western blot and mass spectrometry analyses of anti-Pol II immunoprecipitates obtained from untreated and KCl-treated cells. Surprisingly, rather than observing decreased interactions between Pol II and CPA factors, proteomic analyses revealed that the binding between Integrator subunits and Pol II decreases after stress (Figure 3.1). This was shown by performing anti-Pol II immunoprecipitations followed by mass spectrometry and

anti-Int3 immunoprecipitations followed by western blots using an antibody against Pol II (Figure 3.1). Furthermore, I found that hyperosmotic stress decreases the occupancy of Integrator subunits 11 and 3 on DNA. Therefore, I asked whether siRNA depletion of Int11, the catalytic subunit of the complex, is sufficient to induce DoGs. TT-TL-seq of siRNA-treated samples revealed that hundreds of DoGs are induced after functional Int11 is depleted from cells (Figure 3.1). Moreover, TT-TL-seq results reveal that expression of wild-type Int11 reduces the number of DoGs induced in cells where endogenous Int11 has been knocked down compared to no rescue. Additionally, I observe an increase in the number of DoGs induced upon expression of the E203Q mutant Int11, arguing that the nuclease activity of Int11 is important for repressing DoG production under homeostasis. These results highlight the importance of the Integrator complex in the regulation of mRNA production beyond the promoter proximal pause site (Elrod et al. 2019). Perhaps, cells rely on Integrator to mediate a mechanism of transcription termination at a subset of protein-coding genes that is alternative to the canonical mechanism orchestrated by the CPA machinery. This might explain why readthrough transcription is not observed downstream of all protein-coding genes. However, the ChIP-seq experiments presented here did not reveal preferential loss of Integrator subunits on DoG-producing genes (Figures 3.3 & 3.4). Therefore, it is possible that the genome-wide loss of Integrator upon hyperosmotic stress plays an indirect role in regulating DoG biogenesis. More work investigating the role of Integrator binding to mRNA 3'-ends is imperative.

The results presented in this chapter reveal that as many as 25% of genes that produce DoGs upon KCl-treatment also produce readthrough transcripts upon depletion

of functional Int11 (Figures 3.5E & 3.6G). This partial overlap between Int11-depletion-dependent readthrough transcripts and KCl-induced DoGs suggests that the decreased binding of Integrator subunits to Pol II during the stress response contributes to DoG biogenesis. I asked whether the effects of Int11 knockdown and hyperosmotic stress on DoG induction are additive by exposing HEK-293T cells to KCl after 48 or 72 hours of silInt11 transfection, but did not observe such effects on DoG production upon combined treatment (data not shown). However, these results were inconclusive due to the different extents of DoG induction observed in each context (Figure 3.6H) as well as by the temporal differences inherent to the experimental protocols. Because of the changes in the transcriptional landscape upon hyperosmotic stress (Figures 2.2 & 2.3), it is unlikely that depletion of a single protein is enough to recapitulate the complexity of this response. Hyperosmotic stress induces widespread transcriptional repression to an extent that is not recapitulated by depletion of functional Int11. Additionally, my results reveal changes in Pol II binding patterns to DNA and to other proteins after hyperosmotic stress (Figures 2.4 & 3.1). The full extent of DoG induction upon cellular stress might also reflect a stress-induced increase in the processivity of elongating Pol II molecules, which would result in readthrough transcription (Fong et al. 2015).

In agreement with the important roles of Xrn2 and CPSF73 in mediating the termination of protein-coding gene transcripts (Proudfoot. 2016), readthrough transcription is observed upon depletion of these proteins (Fong et al. 2015; Eaton et al. 2018; Eaton et al. 2020). However, unlike stress-induced readthrough transcription, the effects of depleting these proteins were not limited to a subset of genes. Anti-Pol II immunoprecipitation of chromatin fragments coupled with mass spectrometry and

western blots demonstrated that the overall binding of Pol II and important termination factors, including Xrn2 and CPA factors, is not affected by hyperosmotic stress. Nevertheless, these data do not exclude the possibility that important termination factors are redistributed along the genome after hyperosmotic stress. Anti-CPSF6 ChIP-seq from cells treated with NaCl to induce hyperosmotic stress revealed decreased binding of this subunit at certain genes (Jalihal et al. 2020); yet it is unclear whether this was accompanied by a decrease in Pol II occupancy at these loci. Moreover, anti-CPSF73 ChIP-qPCR data from heat-shocked cells suggest a decrease of this factor at activated DoG-producing genes (Cardiello et al. 2018). It is possible that at genes activated by stress, disruption of the stoichiometry between Pol II and termination factors contributes to DoG production. However, such a model would not explain readthrough transcription at genes that retain comparable expression levels or are transcriptionally repressed, which comprise the majority of DoG-producing genes upon hyperosmotic stress (Figure 2.4B). These observations emphasize the need to further explore the features of DoG-producing genes that might make them susceptible to failed termination despite the presence of important termination factors on chromatin.

The work discussed in this chapter provides insights into the proteomic landscape that accompanies DoG biogenesis after hyperosmotic stress and reveals a role for the Integrator complex in DoG production. However, the Integrator complex does not seem to be the only agent involved in the biogenesis of readthrough transcripts upon stress. Additional work assessing the distribution of other important termination factors on DNA after stress should further elucidate how cells selectively induce DoGs from a subset of genes.

Chapter 4

Materials and methods

Data and code availability

All high throughput sequencing datasets generated for chapters 2 and 3 have been deposited to the Gene Expression Omnibus database and can be found as a super series under the following accession number: GSE152063.

Experimental model and subject details

HEK-293T cells were cultured in DMEM supplemented with 10% fetal bovine serum, 1% penicillin and streptomycin, and 1% L-glutamine. SK-N-BE(2)C cells were cultured in a 1:1 mix of DMEM and F-12K media supplemented with 10% fetal bovine serum, 1% penicillin and streptomycin, and 1% L-glutamine. Stable HEK-293T cell lines expressing WT or E203Q mutant Int11 were supplemented using 2 $\mu\text{g}/\text{mL}$ of puromycin, as previously described (Xie et al. 2015). Hyperosmotic stress inductions were performed using 80 mM KCl for 1 hour, as described in Vilborg et al. 2015.

Method details

RNA preparations and RT-qPCRs: RNA extractions were performed using TRIzol according to the manufacturer's instructions. After treating samples with RQ1 DNase (Promega), RNA was recovered through PCA extraction followed by ethanol precipitation. cDNA was made from 2 μg RNA using Super Script III and random primers (Invitrogen). Samples were then diluted 1:10 and 2 μL of diluted cDNA were used for RT-qPCR along

with 5 μL of iTaq universal SYBR mix (Bio-Rad), 1.5 μL of primer mix (1.5 μM of forward and 1.5 μM of reverse) and 1.5 μL of water. Plates were run on a Bio-Rad CFX384 machine. Results were analyzed using the comparative C_T method (Schmittgen and Livak, 2008). Primers designed for this study were subjected to a primer efficiency test and amplicons were run on agarose gels to ensure that a single band was observed for each pair.

High throughput sequencing: *TT-TimeLapse*: Samples were fed 1 mM $s^4\text{U}$ (Sigma) during the final 5 minutes of incubation with either KCl or siRNAs. After extracting RNA from HEK-293T cells, samples were spiked with 4% $s^4\text{U}$ -labeled RNA from *Drosophila* S2 cells. TT-TL-seq was performed as described in Schofield et al. 2018. Sequencing was done by the Yale Center for Genomic Analysis (YCGA).

ChIP-seq: 2.5×10^7 HEK-293T cells were plated in 15 cm dishes (5×10^6 per dish) and incubated at 37°C overnight. For anti-Int3 ChIP-seq experiments, 3×10^7 HEK-293T cells were used. Cells were crosslinked during the last 10 minutes of stress induction using 1% formaldehyde and then washed twice with 1x PBS. Chromatin immunoprecipitation was performed as described in Bieberstein et al. 2014 using 30 μL of anti-Pol II antibody (Cell Signaling D8L4Y), 25 μL of anti-Int11 antibody (Abcam ab75276) or 30 μL of anti-Int3 antibody (Bethyl A302-050A) and 180 μL of Protein A beads. The YCGA prepared libraries and sequenced the samples on the NovaSeq platform. I note that the generation of Int11 ChIP-seq datasets was particularly challenging. I suspect that this was due to the biological properties of this protein as well as the reduced interactions between Integrator

subunits and Pol II upon hyperosmotic stress. Subsequently, I performed anti-Int3 ChIP-seq experiments and obtained agreeing results.

Bioinformatics: *TT-TimeLapse-seq*: TT-TL-seq experiments were performed in collaboration with Joshua Zimmer. Reads from HEK-293T cells containing spiked-in RNA from *Drosophila* S2 cells were mapped to a combined hg38 and dm6 genome using hisat2 version 2.1.0 (Kim et al. 2015). Bam files for mapped reads were created, sorted and indexed using samtools version 1.9 (Li et al. 2009). U-to-C mutation were identified as described in Schofield et al. (2018). Normalization factors for each sample were calculated in edgeR (Robinson et al. 2010) using read counts mapped to the dm6 genome obtained from HTSeq (Anders et al. 2015).

Sample	Spike-in normalization factor
siC_rep1	2.65151515
KCl_rep1	0.71138211
silnt11_rep1	2.08333333
siC_rep2	1.94444444
KCl_rep2	0.63405797
silnt11_rep2	1.26811594
WT	2.65151515
mut	1.82291667

Gene counts for reads aligning to the hg38 genome were generated using bedtools multicov version 2.26.0 (Quinlan et al. 2010). Read-in values used to define clean genes were generated using ARTDeco (Roth et al. 2020). DoGs were identified using

DoGFinder (Wiesel et al. 2018). Differential expression analysis of normalized read counts in gene bodies and in DoG regions was done using DESeq2 (Love et al. 2014). The list of clean genes is comprised by genes that were expressed by more than 100 read counts after normalization to spike-in, had a read in value ≤ -1 and did not overlap with readthrough regions from neighboring genes on either strand according to bedtools intersect analysis (Quinlan et al. 2010). EnrichR was used for gene ontology analysis of biological enrichments (Chen et al. 2013; Kuleshov et al. 2016). Visualization of the generated data was achieved using ggplot2 (Wickham 2016) and graphPad prism. Eulerr was used to generate proportional Venn diagrams. Normalized tracks were visualized using IGV.

ChIP-seq: ChIP-seq samples were mapped to the hg38 human genome using bowtie2 (Langmead et al. 2009). Bam files and gene counts were generated as described above. Counts were normalized to input using edgeR normalizeChIPtoInput function (Robinson et al. 2010). MACS2 version 2.1.1 (Zhang et al. 2008) was used to call peaks. Volcano plots were generated using ggplot2 on \log_2 fold changes and adjusted p-values calculated with DESeq2. Normalized bigwig files and meta plots were created using deepTools version 3.3.0 (Ramírez et al. 2016). Tracks were visualized using IGV.

Cell fractionation and immunoprecipitations: For cellular fractionation, 4.8×10^6 cells were plated in three 15 cm dishes and incubated overnight at 37°C. After KCl treatment, cells were washed twice with 1x PBS, scraped on ice and pelleted by centrifugation (1400rpm at 4°C for 5 minutes). Pellets were resuspended using hypotonic lysis buffer for 5 minutes (Nojima et al. 2016). Nuclei were collected by centrifugation (1400rpm at 4°C for 5

minutes) and the cytoplasmic fraction (supernatant) was discarded. Nuclei were incubated in NUN1 and NUN2 (Wuarin and Schibler 1994) on ice for 15 minutes with intermittent vortexing. Samples were then spun down to pellet the chromatin fraction. Chromatin fragments were generated by incubating the chromatin pellets with 2 μ L of micrococcal nuclease for 5 minutes at 37°C, at 1400 rpm in a thermomixer (MNase-digested samples) or by incubating the pellets with 25U of benzonase for 45 minutes (Benzonase-digested samples). After digestion, the samples were clarified (16,000 g at 4°C for 5 minutes) and the supernatant was transferred into a new 1.5 mL tube.

For immunoprecipitation, 10 μ g of anti-Pol II antibody (MABI0601) or 3 μ g of anti-Int3 (Bethyl A302-050A) antibody were incubated with 25 μ L of magnetic anti-mouse beads (NEB) or Protein A beads, respectively, overnight. The following day, the antibody-conjugated beads were washed twice with 1 mL of NET-2 buffer and resuspended in 25 μ L of NET-2 buffer. Chromatin fragments were pre-cleared by incubating them with the IgG-conjugated beads for 3 hours in a rotator at 4°C. The beads were then collected using a magnetic rack and the supernatant was transferred to a 1.5 mL tube containing the anti-Pol II conjugated beads. The samples were incubated for an additional 2 hours in the rotator at 4°C. After collecting the beads using a magnetic rack, the supernatant was discarded and the beads were washed four times using 1 mL of NET-2 buffer. The final wash was done using 1x PBS and the samples were resuspended using 8 μ L of 10 mM Tris-HCL pH 8.5.

Proteomics: Mass spectrometry was performed by Dr. Maria Apostolidi. Briefly, 3-4 μ g of peptides obtained from anti-Pol II immunoprecipitates were subjected to LC-MS/MS.

Mass spectra were analyzed using Maxquant version 1.5.1.2 (Cox et al. 2008). For additional details, see Rosa-Mercado et al. 2021.

siRNA knockdowns: For knockdown experiments, 7.5×10^4 HEK-293T cells were seeded in six well plates and incubated at 37°C overnight. After 24 hours, the cells were transfected with either 50 nM of scrambled siRNA control (siC), siRNA against Int11 (silInt11) (Xie et al. 2015) or siRNA against Int3 (silInt3) (Skaar et al. 2015) using Lipofectamine RNAiMax according to manufacturer's instructions. Cells were incubated at 37°C for 72 hours after which they were washed with 1x PBS, scraped and pelleted by centrifugation (1400 rpm at 4°C for 5 minutes). After discarding the supernatant, cells were resuspended in 150 μ L of NET-2 buffer. 50 μ L were aliquoted to verify knockdown efficiency through western blots and the rest of the sample was lysed using TRIzol. siRNA experiments for TT-TL experiments were performed by transfecting 1.5×10^5 cells for 48 hours instead of 72 hours.

Western blots: For western blots of whole cell lysates, samples were sonicated using 10 cycles of 30 seconds on and 30 seconds off in a Diagenode Bioruptor Pico sonicator (Withers et al. 2018) and digested with micrococcal nuclease for 45 minutes at 37°C. Samples were run at an increasing voltage (up to 120 V) in NuPAGE 4-12% BisTris gels in 1x MOPS buffer (Invitrogen). Proteins were transferred overnight to a 0.45 μ m nitrocellulose membrane at 30 V. Membranes were blocked for 1 hour in 5% milk in 1x PBST. Detection of total Pol II and Tyr1-P containing Pol II molecules was achieved using antibodies from MBL (MABI0601, discontinued) and Active motif (61383) respectively, at a 1:1000 dilution. Antibodies against Ints3 (Bethyl A302-050A), Ints11 (Abcam ab75276),

Xrn2 (A301-101) and CPSF160 (A301-580A) were used at a 1:800 dilution. Anti-GAPDH antibody (Cell Signaling 14C10) was used at 1:2000, while the antibody against U1-70K (Kastner et al. 1992; Tarn and Steitz. 1994) was used at a 1:500 dilution. All secondary antibodies were used at a 1:2000 dilution of 1x PBST, 3% milk.

Quantification and statistical analysis

One sample t-tests and paired t-tests were performed on the fold changes of western blot band intensities and on RT-qPCR data, respectively, using graphPad prism. Differential expression analysis for TT-TL-seq and ChIP-seq datasets were obtained using DESeq2.

Materials table

REAGENT	SOURCE	IDENTIFIER
Antibodies		
CPSF1	Bethyl Laboratories	A301-580A
GAPDH	Cell Signaling	14C10
H3	Abcam	ab1791
Int3	Bethyl Laboratories	A302-050A
Int11	Abcam	ab75276
RNA Pol II CTD	MBL	MABI0601
RNA Pol II (N-terminus)	Cell Signaling	D8L4Y
RNA Pol II Tyr1-P CTD	Active Motif	61383
U1-70K	Kastner et al. 1992; Tarn and Steitz. 1994	N/A
Xrn2	Bethyl Laboratories	A301-101

Chemicals, peptides, and recombinant proteins		
Dynabeads Protein A	Invitrogen	1001D
4-thiouridine	Sigma	T4509
Goat anti-mouse beads	NEB	S1431S
Lipofectamine RNAimax	Invitrogen	13778-150
Micrococcal nuclease	NEB	M0247S
RQ1 DNase	Promega	M610A
Superscript III	Invitrogen	56575
Deposited data		
Raw and analyzed TT-TL-seq data	This paper	GSE152059
Raw and analyzed anti-Pol II ChIP-seq data	This paper	GSE152062
Raw and analyzed anti-Int3 and anti-Int11 data	This paper	GSE159190
Experimental models: Cell lines		
HEK-293T	Steitz lab	N/A
HEK-293T 2x FLAG-Int11 WT	Xie et al. 2015	N/A
HEK-293T 2x FLAG-Int11 E203Q	Xie et al. 2015	N/A
<i>Drosophila</i> S2 cells	Simon lab	N/A
SK-N-BE(2)C	Steitz lab	N/A
Oligonucleotides		
Random primers	Invitrogen	58875

Software and algorithms		
ARTDeco	Roth et al. 2020	N/A
bedtools	Quinlan et al. 2010	N/A
Bowtie2	Langmead et al. 2009	N/A
Cutadapt	Martin 2011	N/A
DESeq2	Love et al. 2014	N/A
DoGFinder	Wiesel et al. 2018	N/A
edger	Robinson et al. 2010	N/A
enrichr	Chen et al. 2013; Kuleshov et al. 2016	N/A
eulerr	Larsson J (2020)	https://github.com/jolars/eulerr
FastUniq	Xu et al. 2012	N/A
ggplot2	Wickham. 2016	N/A
Hisat2	Kim et al. 2015	N/A
HTSeq	Anders et al. 2015	N/A
MACS2	Zhang et al. 2008	N/A
Prism	GraphPad	N/A
samtools	Li et al. 2009	N/A

Oligos used in this study.

RT-qPCR		
Name	Oligo	Reference
doSERBP1 F	GCCTTATGTTACCCAGCCTCT	Vilborg et al. 2015

doSERBP1 R	AGCAGGAACTTTGGCAGTGT	Vilborg et al. 2015
doRAB11A F	TTCTCCCCTAGAAGGCTGTGT	Vilborg et al. 2015
doRAB11A R	AGCAAGAGGTGGAATGGCTA	Vilborg et al. 2015
doINTS6 F	CCGATACACTGGGCTTGCA	N/A
doINTS6 R	ACCCTAAGCCACCAACAGTC	N/A
doSNRPB F	CAGGGTTCACAGCAGCTATTAT	N/A
doSNRPB R	GGTGCTCGAGAAGAGTGTATTC	N/A
doSERTAD1 F	ACATGCGCTTTGGGGTAAT	N/A
doSERTAD1 R	CACATGCCACTTGAACCCAC	N/A
doDDX18 F	CCCATGTGAAAACAGAGCAA	Vilborg et al. 2015
doDDX18 R	GTTTGCCTGTACGGTCCATT	Vilborg et al. 2015
DDX18 mRNA F	AAGCAGCAATGCAGTCTTCC	Vilborg et al. 2015
DDX18 mRNA R	TCTCGGCACTTTCTTCTTCAG	Vilborg et al. 2015
pre-DDX18 F	GATAGCCAAGGCACAGTCCA	N/A
pre-DDX18 R	TTTCTTCCCACGCCAATCACT	N/A
doOPA1 F	GTTCCCTGAACACTCTGCGT	N/A
doOPA1 R	GGCGATATGGGTTTGTGAAGC	N/A
OPA1 mRNA F	TTCCTGGGTCATTCTGGAC	N/A

OPA1 mRNA R	TTTCTGCTATCCAGGCCACAG	N/A
pre-OPA1 F	ATTACGGGGAGCAATCCACTG	N/A
pre-OPA1 R	TGGAAAAGGGATTGTCGGCT	N/A
doNFX1 F	GTGAAAGCACAGGCTTCACG	N/A
doNFX1 R	GCACATCTGACACACAGGGA	N/A
NFX1 mRNA F	CAAACCTGCGCTAGAGTCCA	N/A
NFX1 mRNA R	GGGGATGTTGCTCCGAACT	N/A
pre-NFX1 F	AGCCAGGGACTGTAACACCA	N/A
pre-NFX1 R	AGATGGCTCTAGCCGTTGC	N/A
GAPDH mRNA F	CTGCACCACCAACTGCTTAG	Vilborg et al. 2015
GAPDH mRNA R	GTCTTCTGGGTGGCAGTGAT	Vilborg et al. 2015
GUSB mRNA F	AAACGATTGCAGGGTTTCAC	Vilborg et al. 2015
GUSB mRNA R	TCCAAATGAGCTCTCCAACC	Vilborg et al. 2015
Unprocessed U2 F	GGAGCAGGGAGATGGAATAGGAGCTT	Skaar et al. 2015
Unprocessed U2 R	CAGGACTCGTGCAAGCCGCC	Skaar et al. 2015

Northern blot probes		
Target	Sequence	Reference

U2	CGTTCCTGGAGGTACTGCAA	Giacometti et al. 2017
U6	GCAGGGGCCATGCTAATCTTCTCTGTATCG	Pfeffer et al. 2004
Ile-tRNA	GGTGAGGCTCGAACTCACAACCTCGGC	Eckwahl et al. 2015

siRNAs		
Name	Sequence	Reference
siC	rArArGrCrGrArUrArCrCrUrCrGrUrGrUrGrUrGrArUrU	Xie et al. 2015
silnt11	rUrCrGrArArGrGrCrCrUrUrGrArUrGrUrGrCrUrUrU	Xie et al. 2015
silnt3	rGrArUrGrArGrArGrUrUrGrCrUrArUrGrArCrArUrU	Skaar et al. 2015

Chapter 5

Discussion and future directions

Summary:

The findings presented in this dissertation contribute to our understanding of the biogenesis of readthrough transcripts induced by hyperosmotic stress, implicating the Integrator complex in DoG production. However, the exact mechanistic details regarding how Integrator contributes to DoG induction upon stress remain unclear. To verify if post-translational modifications that favor the interactions between Pol II and Integrator are affected by hyperosmotic stress, I performed ChIP-seq experiments against Pol II phosphorylated in tyrosine 1 or serine 7 residues of the C-terminal domain (CTD) of Rpb1, the largest subunit of the complex. Preliminary results reveal no difference in occupancy of Pol II molecules phosphorylated at tyrosine 1 or serine 7 of the CTD at genes that are not repressed by hyperosmotic stress when comparing stressed cells to untreated cells. More work needs to be done to obtain a full picture of the mechanisms of DoG biogenesis upon stress.

Hyperosmotic stress induces the transcriptional repression of many genes. The contribution of DoG production to this transcriptional response remains elusive. Furthermore, the dynamics of recovery from stress-induced transcriptional regulation compared to the recovery of efficient transcription termination are unknown. My preliminary results suggest that DoGs are detected among nascent transcripts for up to an hour after stress removal, while the transcriptional response of their parent genes begins to recover within fifteen minutes of stress removal. Future experiments will assess the dynamics of DoG recovery relative to the transcriptional response genome-wide.

Discovery of additional factors that contribute to DoG production upon stress

Termination factors and histone modifications have been found to play a role in DoG production (Chapter 1). My results implicate the Integrator complex as a contributor to DoG biogenesis (Chapter 3; Figure 5.1). However, my results do not provide information about the exact mechanism by which Integrator regulates DoG biogenesis. More work is needed to determine whether the effect of the Integrator complex on readthrough transcripts produced from protein-coding genes is direct or indirect.

Phosphorylation of the C-terminal domain of Rpb1, the largest subunit of Pol II, regulates many steps of transcription including initiation, elongation, splicing and termination (reviewed in Harlen and Churchman. 2017). The Integrator complex interacts with RNA Pol II molecules that are phosphorylated in tyrosine 1 (Tyr-1) or serine-7 (Ser-7) within the heptapeptide repeats of the Pol II CTD (Shah et al. 2018; Egloff et al. 2007). On the other hand, phosphorylation of serine 5 (Ser-5) of the CTD is unfavorable to the interactions between Pol II and Integrator (Egloff et al. 2007). To investigate if hyperosmotic stress induces changes in the phosphorylation of Tyr-1 or Ser-7 of the Pol II CTD, I performed ChIP-seq experiments using antibodies that specifically identify Pol II molecules containing these post-translational modifications. My preliminary results indicate that hyperosmotic stress does not disrupt the phosphorylation of Tyr-1 or Ser-7 of the Pol II CTD as their levels remain unaffected for genes that do not experience transcriptional repression according to the TT-TL-seq data discussed in Chapter 2 (Figure 5.2). Experiments to investigate the distribution of other phosphorylated residues of the Pol II CTD in cells exposed to hyperosmotic stress have proven challenging due to difficulties in obtaining sufficient DNA for preparation of ChIP-seq libraries. This is likely

due at least in part to the decreased binding of Pol II molecules along the bodies of repressed genes in cells that have been exposed to hyperosmotic stress (Figure 2.5).

Other groups have suggested that DoGs result from dysregulated occupancy of CPA factors (Cardiello et al. 2018; Jalihal et al. 2020). Yet, how this dysregulated occupancy specifically affects DoG-producing genes is not understood. Therefore, more work needs to be done in order to obtain a full picture of the mechanisms responsible for DoG synthesis. Experiments that assess the binding of termination factors to DoG-producing genes or to DoG transcripts may provide additional information about the interactions that facilitate cleavage of nascent RNAs in cells experiencing stress.

Assessing the dynamics of transcriptional regulation and DoG production upon hyperosmotic stress and recovery

DoGs are induced within minutes after cells are exposed to hyperosmotic stress (Vilborg et al. 2017). To gain insight into the dynamics of transcriptional regulation and DoG induction upon stress, I performed RT-qPCR analyses on RNA obtained from cells exposed to hyperosmotic stress for different times using primers against the pre-mRNAs of DoG-producing genes that are either activated, repressed or unchanged upon stress. Preliminary results reveal that the onset of transcriptional regulation of DoG-producing genes coincides with DoG induction (Figure 5.3 A & B). This suggests that transcriptional output and transcription termination are regulated within minutes of stress exposure, thereby contributing to the first steps of the cellular stress response.

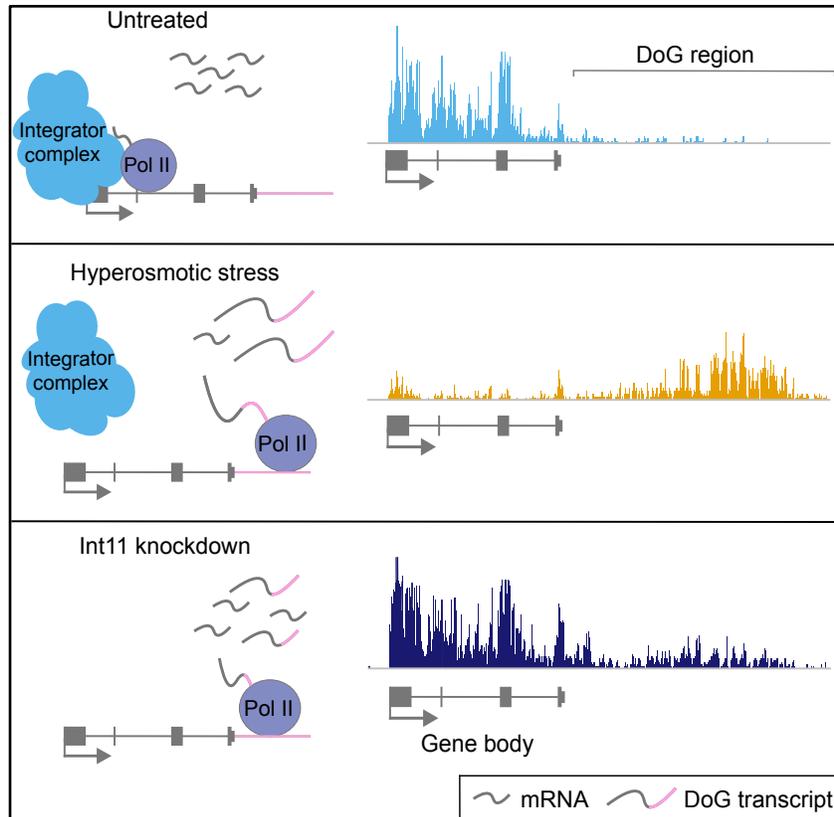


Figure 5.1: Cartoon depiction of DoG induction upon hyperosmotic stress or Int11 knockdown. Hyperosmotic stress causes widespread transcriptional repression in human cells, yet downstream-of-gene transcripts (DoGs) arise regardless of the transcriptional response of their upstream genes. Interactions between Pol II and Integrator are disrupted by hypertonicity. Knock down of the Integrator nuclease (Int11) leads to DoG production.

Cells are able to recover efficient transcription termination upon stress removal: after 3.5 hours of stress removal, DoGs are no longer detected among total RNAs from cells that had been treated with mild conditions of hyperosmotic stress for one hour (Vilborg et al. 2017). To investigate if the DoG transcripts detected in RNA obtained from cells allowed to recover from stress for shorter periods of time represent nascent transcripts, I exposed cells to hyperosmotic stress for one hour, allowed them to recover for different time periods, and incubated cells with s⁴U during the last 5 minutes of recovery. Preliminary results show that certain DoGs are detected among nascent transcripts for up to an hour after stress removal while others return to base levels minutes after stress removal (Figure 5.3C). RT-qPCR analyses performed using primers against the first or second intron of the corresponding DoG-producing genes revealed that pre-mRNA from genes that are differentially regulated upon stress return to approximately basal levels within fifteen minutes of stress removal (Figure 5.3D).

The discrepancy in the dynamics of recovery between DoG transcripts suggests that the decay of these RNAs might be regulated by different pathways. In collaboration with Professor Matthew Simon's lab, I plan to assess the pattern of DoG recovery in human cells exposed to hyperosmotic stress genome-wide. TimeLapse-seq data obtained from cells allowed to recover from stress will also allow us to investigate whether or not recovery from stress induces changes in the half-lives of DoGs. These experiments will shed light into the dynamics of DoG induction and decay in human cells.

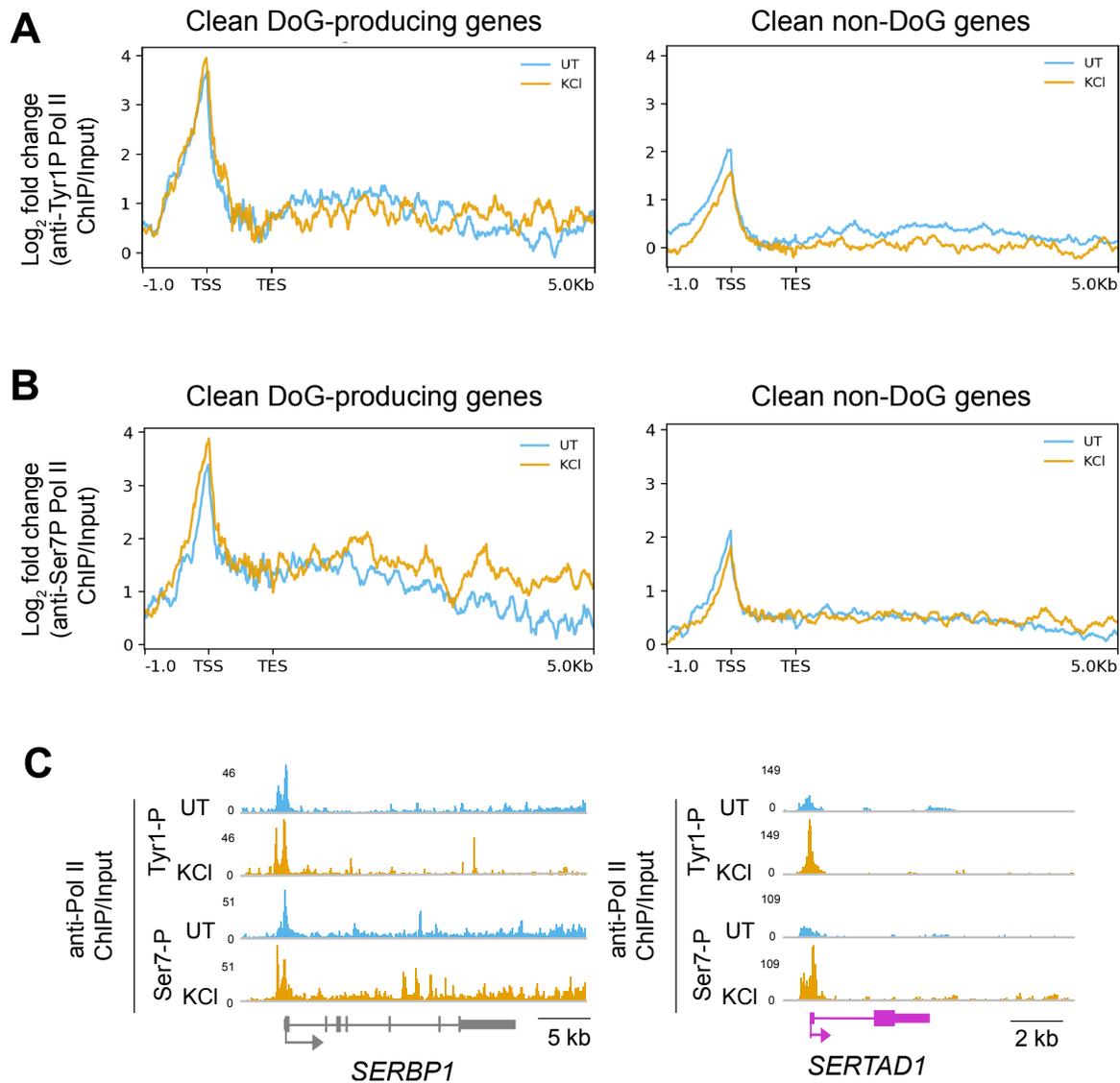


Figure 5.2: Phosphorylation of Pol II CTD residues tyrosine-1 and serine-7 does not decrease after hyperosmotic stress. A, B) Meta plots showing the log₂ fold change of ChIP/Input for A) anti-Tyr1-P and B) anti-Ser7-P Pol II ChIP-seq data for DoG-producing genes (n=71) or non-DoG genes (n=467) that are not repressed upon hyperosmotic stress according to TT-TL-seq. C) Browser images of *SERBP1*, which retains comparable levels of expression in KCl-treated cells, and for *SERTAD1*, which is transcriptionally activated upon stress.

Investigating the impact of DoG transcription on read-in gene transcript levels

DoG production leads to the invasive transcription of neighboring loci (Rutkowski et al. 2015; Grosso et al. 2015; Roth et al. 2020; Figure 2.1). However, the impact of this invasive transcription on the mRNA levels of read-in genes is unknown. It is possible that read-in transcription contributes to the transcriptional repression observed upon exposure to different stress conditions by causing transcriptional interference.

Investigating the impact of DoG production on the nascent transcription profiles of read-in genes cannot be achieved by traditional short read sequencing techniques or by RT-qPCR. Instead, this can be done by using targeted long read sequencing and more quantitative approaches that measure the levels of Pol II occupancy near the transcription start sites of read-in genes by examining the associated short, capped transcripts (Nechaev et al. 2010; Zimmer et al. 2021). Additionally, the impact of DoG transcription on the levels of read-in gene transcripts can be assessed by using a dead Cas9 system to target the upstream DoG-producing gene, which would result in lower levels of the corresponding DoG (Vilborg et al. 2015). Levels of the read-in transcript in stressed versus unstressed cells with or without the production of the upstream DoG can then be measured through northern blots. Such approaches will provide a comprehensive understanding on the impact of DoG biogenesis on the accompanying transcriptional landscapes, thereby shedding light on a potential function of these stress-induced readthrough transcripts.

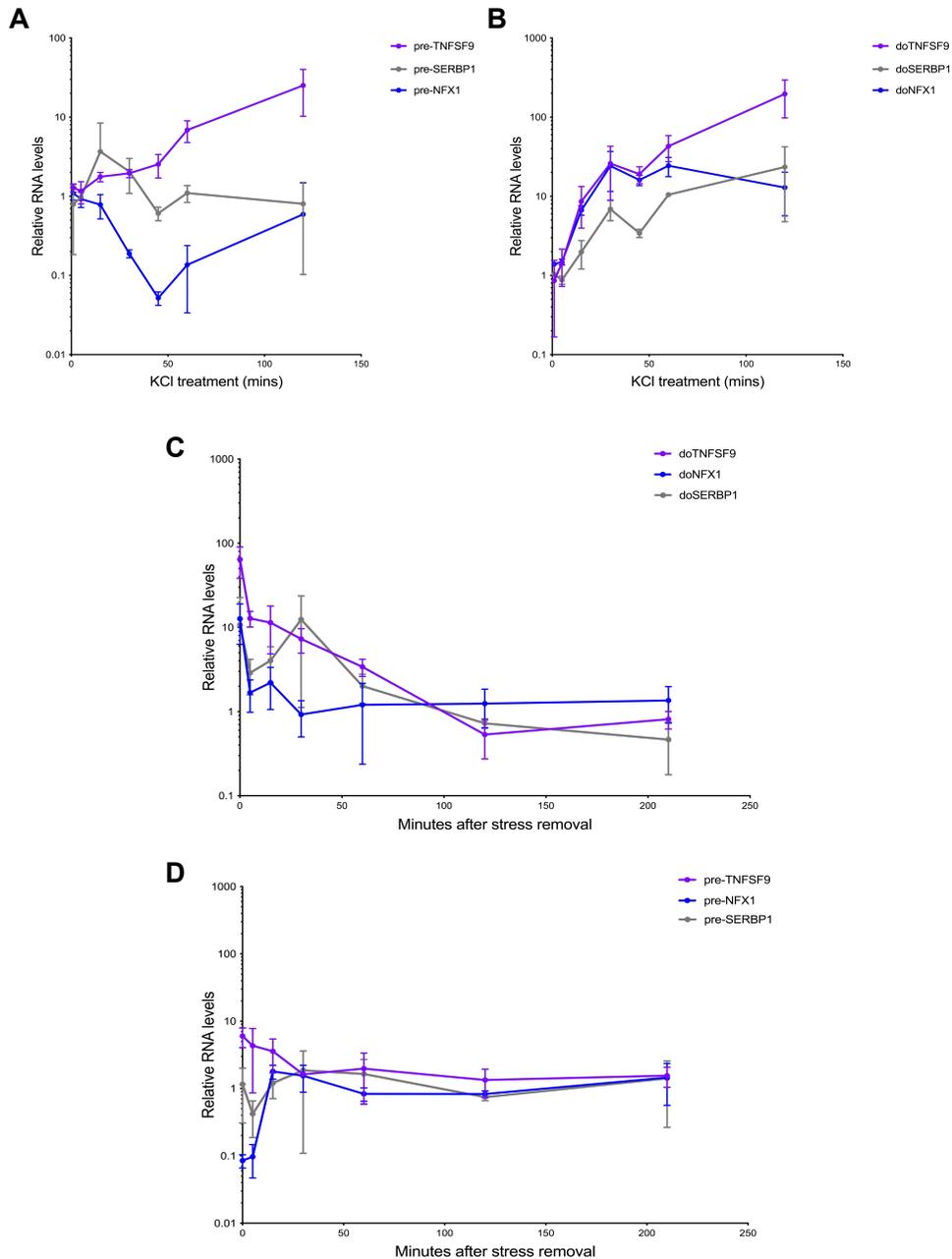


Figure 5.3: Dynamics of transcriptional regulation and DoG induction at DoG-producing genes. A, B) Time course of A) transcriptional regulation of upstream DoG-producing genes or B) DoG induction in HEK-293T cells, both measured by RT-qPCR (n=3 for 5, 15, 45 and 60 minutes; n=2 for 120 minutes). According to the TT-TL-seq data discussed in chapter 2, *TNFSF9* is transcriptionally activated, *SERBP1* retains comparable levels of expression and *NFX1* is repressed upon hyperosmotic stress. C, D) Time course of stress recovery for C) DoG and D) pre-mRNA levels from 3 DoG-producing genes in HEK-293T cells measured by RT-qPCR of s⁴U-enriched RNAs (n=3 for 0, 5, 15, 30, 60, 120 and 210 minutes).

Outstanding questions

1. Are DoGs produced in response to specific signaling pathways activated upon stress?
2. Do cells employ multiple strategies to produce DoGs or do known factors converge into a single pathway?
3. Are there as-yet unrecognized factors involved in DoG biogenesis?
4. What characteristic features make DoG-producing genes more likely to read through than non-DoG genes?
5. How does DoG transcription impact the levels of neighboring, read-in genes?
6. What is the function of DoGs?

References

- Albrecht, T.R., Shevstov, S.P., Wu, Y., Mascibroda, L.G., Peart, Natoya 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, 4241-4255.
- Albrecht, T.R., and Wagner, E.J. (2012). snRNA 3' end formation requires heterodimeric association of integrator subunits. *Mol Cell Biol* 32, 1112-1123.
- Alpert, T., Straube, K., Carrillo Oesterreich, F., Herzelt, L., and Neugebauer, K.M. (2020). Widespread Transcriptional Readthrough Caused by Nab2 Depletion Leads to Chimeric Transcripts with Retained Introns. *Cell Rep* 33, 108324.
- Amat, R., Bottcher, R., Le Dily, F., Vidal, E., Quilez, J., Cuartero, Y., Beato, M., de Nadal, E., and Posas, F. (2019). Rapid reversible changes in compartments and local chromatin organization revealed by hyperosmotic shock. *Genome Res* 29, 18-28.
- Anders S., Pyl P.T., and Huber W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), 166-9.
- Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA Polymerase II. *Cell* 123, 265-276.
- Baillat, D., and Wagner, E.J. (2015). Integrator: surprisingly diverse functions in gene expression. *Trends Biochem Sci* 40, 257-264.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research* 21, 381-395.
- Bauer, D.L.V., Tellier, M., Martinez-Alonso, M., Nojima, T., Proudfoot, N.J., Murphy, S., and Fodor, E. (2018). Influenza virus mounts a two-pronged attack on host RNA Polymerase II transcription. *Cell Rep* 23, 2119-2129 e2113.
- Beagan, J.A., and Phillips-Cremins, J.E. (2020). On the existence and functionality of topologically associating domains. *Nature Genetics* 52, 8-16.
- Beckedorff, F., Blumenthal, E., daSilva, L.F., Aoi, Y., Cingaram, P.R., Yue, J., Zhang, A., Dokaneheifard, S., Valencia, M.G., Gaidosh, G., *et al.* (2020). The human Integrator complex facilitates transcriptional elongation by endonucleolytic cleavage of nascent transcripts. *Cell Rep* 32, 107917.

- Bieberstein, N.I., Straube, K., and Neugebauer, K.M. (2014). Chromatin immunoprecipitation approaches to determine co-transcriptional nature of splicing. *Methods Mol Biol* 1126, 315-323.
- Birkenheuer, C.H., Danko, C.G., and Baines, J.D. (2018). Herpes Simplex Virus 1 Dramatically Alters Loading and Positioning of RNA Polymerase II on Host Genes Early in Infection. *Journal of Virology* 92, e02184-02117.
- Boutz, P.L., Bhutkar, A., and Sharp, P.A. (2015). Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev* 29, 63-80.
- Cardiello, J.F., Goodrich, J.A., and Kugel, J.F. (2018). Heat shock causes a reversible increase in RNA Polymerase II occupancy downstream of mRNA genes, consistent with a global loss in transcriptional termination. *Mol Cell Biol* 38.
- Castillo-Guzman, D., Hartono, S.R., Sanz, L.A., Chédin, F. (2020). SF3B1-targeted splicing inhibition triggers global alterations in transcriptional dynamics and R-loop metabolism. *bioRxiv*.
- Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14, 128
- Chen, K., Hu, Z., Xia, Z., Zhao, D., Li, W., and Tyler, J.K. (2015). The overlooked fact: fundamental need for spike-in control for virtually all genome-wide analyses. *Mol Cell Biol* 36, 662-667.
- Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367-1372.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences.* 107, 21931-21936.
- De Nadal, E., Ammerer, G., and Posas, F. (2011). Controlling gene expression in response to stress. *Nat. Rev. Genet.* 12, 833–845.
- 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. *Mol Cell Biol* 25, 1489-1500.
- Duffy, E.E., Rutenberg-Schoenberg, M., Stark, C.D., Kitchen, R.R., Gerstein, M.B., and Simon, M.D. (2015). Tracking distinct RNA populations using efficient and reversible covalent chemistry. *Mol Cell* 59, 858-866.
- Duns, G., van den Berg, E., van Duivenbode, I., Osinga, J., Hollema, H., Hofstra, R.M., and Kok, K. (2010). Histone methyltransferase gene SETD2 is a novel tumor suppressor gene in clear cell renal cell carcinoma. *Cancer Res* 70, 4287-4291.
- Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., and West, S. (2018). Xrn2 accelerates termination by RNA Polymerase II, which is underpinned by CPSF73 activity. *Genes Dev* 32, 127-139.
- Eaton, J.D., Francis, L., Davidson, L., and West, S. (2020). A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev* 34, 132-145.
- Eckwahl, M.J., Sim, S., Smith, D., Telesnitsky, A., and Wolin, S.L. (2015). A retrovirus packages nascent host noncoding RNAs from a novel surveillance pathway. *Genes Dev* 29, 646-657.
- Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* 318, 1777-1779.
- Elrod, N.D., Henriques, T., Huang, K.L., Tatomer, D.C., Wilusz, J.E., Wagner, E.J., and Adelman, K. (2019). The Integrator complex attenuates promoter-proximal transcription at protein-coding genes. *Mol Cell* 76, 738-752.e737.
- Finan, J.D., and Guilak, F. (2010). The effects of osmotic stress on the structure and function of the cell nucleus. *J. Cell. Biochem.* 109, 460–467.
- Finan, J.D., Leddy, H.A., and Guilak, F. (2011). Osmotic stress alters chromatin condensation and nucleocytoplasmic transport. *Biochem. Biophys. Res. Commun.* 408, 230–235.
- Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M.A., Sheridan, R.M., Nguyen, T., Karp, S., and Bentley, D.L. (2015). Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA Polymerase II termination suggest widespread kinetic competition. *Mol Cell* 60, 256-267.

- Forero, A., Ozarkar, S., Li, H., Lee, C.H., Hemann, E.A., Nadjombati, M.S., Hendricks, M.R., So, L., Green, R., Roy, C.N., *et al.* (2019). Differential Activation of the Transcription Factor IRF1 Underlies the Distinct Immune Responses Elicited by Type I and Type III Interferons. *Immunity* 51, 451-464.e456.
- Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J.M., Wagner, E.J., Lazar, M.A., Shilatifard, A., and Shiekhattar, R. (2014). Integrator regulates transcriptional initiation and pause release following activation. *Mol Cell* 56, 128-139.
- Gates, L.A., Foulds, C.E., and O'Malley, B.W. (2017). Histone Marks in the 'Driver's Seat': Functional Roles in Steering the Transcription Cycle. *Trends Biochem Sci* 42, 977-989.
- Giacometti, S., Benbahouche, N.E.H., Domanski, M., Robert, M.C., Meola, N., Lubas, M., Bukenborg, J., Andersen, J.S., Schulze, W.M., Verheggen, C., *et al.* (2017). Mutually Exclusive CBC-Containing Complexes Contribute to RNA Fate. *Cell Rep* 18, 2635-2650.
- Giaimo, B.D., Ferrante, F., Herchenröther, A., Hake, S.B., and Borggreffe, T. (2019). The histone variant H2A.Z in gene regulation. *Epigenetics Chromatin* 12, 37.
- Greer, P.L., and Greenberg, M.E. (2008). From Synapse to Nucleus: Calcium-Dependent Gene Transcription in the Control of Synapse Development and Function. *Neuron* 59, 846-860.
- Grosso, A.R., Leite, A.P., Carvalho, S., Matos, M.R., Martins, F.B., Vitor, A.C., Desterro, J.M., Carmo-Fonseca, M., and de Almeida, S.F. (2015). Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. *Elife* 4.
- Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II interactomes reveal distinct and varied roles for each phospho-CTD residue. *Cell Rep* 15, 2147-2158.
- Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol* 18, 263-273.
- Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., *et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39, 311-318.

- Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., White, K.M., Albrecht, R.A., Pache, L., *et al.* (2018). Transcription elongation can affect genome 3D structure. *Cell* 174, 1522-1536.e1522.
- Hennig, T., Michalski, M., Rutkowski, A.J., Djakovic, L., Whisnant, A.W., Friedl, M.S., Jha, B.A., Baptista, M.A.P., L'Hernault, A., Erhard, F., *et al.* (2018). HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes. *PLoS Pathog* 14, e1006954.
- Herzel, L., Straube, K., and Neugebauer, K.M. (2018). Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Res* 28, 1008-1019.
- Himanen, S.V., and Sistonen, L. (2019). New insights into transcriptional reprogramming during cellular stress. *J Cell Sci* 132.
- Huang, C., and Zhu, B. (2018). Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. *Biophys Rep* 4, 170-177.
- Jalihal, A.P., Pitchiaya, S., Xiao, L., Bawa, P., Jiang, X., Bedi, K., Parolia, A., Cieslik, M., Ljungman, M., Chinnaiyan, A.M., *et al.* (2020). Multivalent proteins rapidly and reversibly phase-separate upon osmotic cell volume change. *bioRxiv*, 748293.
- Kastner, B., Kornstadt, U., Bach, M., and Luhrmann, R. (1992). Structure of the small nuclear RNP particle U1: identification of the two structural protuberances with RNP-antigens A and 70K. *J Cell Biol* 116, 839-849.
- Khosla, S., Farr, J.N., Tchkonja, T., and Kirkland, J.L. (2020). The role of cellular senescence in ageing and endocrine disease. *Nat Rev Endocrinol* 16, 263-275.
- Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12, 357-360.
- Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., *et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182-187.

- Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., *et al.* (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 44, W90-97.
- Lai, F., Gardini, A., Zhang, A., and Shiekhata, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. *Nature* 525, 399-403.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078-2079.
- Li, Y.-J., Fu, X.-H., Liu, D.-P., and Liang, C.-C. (2004). Opening the chromatin for transcription. *The International Journal of Biochemistry & Cell Biology* 36, 1411-1423.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289-293.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550.
- Lovén, J., Orlando, D.A., Sigova, A.A., Lin, C.Y., Rahl, P.B., Burge, C.B., Levens, D.L., Lee, T.I., and Young, R.A. (2012). Revisiting global gene expression analysis. *Cell* 151, 476-482.
- Lykke-Andersen, S., Žumer, K., Molska, E., Rouvière, J.O., Wu, G., Demel, C., Schwalb, B., Schmid, M., Cramer, P., and Jensen, T.H. (2021). Integrator is a genome-wide attenuator of non-productive transcription. *Mol Cell*.
- Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G., and Lis, J.T. (2016). Mammalian heat shock response and mechanisms underlying its genome-wide transcriptional regulation. *Mol Cell* 62, 63-78.
- Mak, S.K., and Kültz, D. (2004). Gadd45 proteins induce G2/M arrest and modulate apoptosis in kidney cells exposed to hyperosmotic stress. *J Biol Chem* 279, 39075-39084.

- Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* *444*, 953-956.
- Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, *17*, 10-12.
- Mazo, A., Hodgson, J.W., Petruk, S., Sedkov, Y., and Brock, H.W. (2007). Transcriptional interference: an unexpected layer of complexity in gene regulation. *J Cell Sci* *120*, 2755-2761.
- Muniz, L., Deb, M.K., Aguirrebengoa, M., Lazorthes, S., Trouche, D., and Nicolas, E. (2017). Control of gene expression in senescence through transcriptional read-through of convergent protein-coding genes. *Cell Rep* *21*, 2433-2446.
- Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* *327*, 335-338.
- Nemeroff, M.E., Barabino, S.M., Li, Y., Keller, W., and Krug, R.M. (1998). Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3'end formation of cellular pre-mRNAs. *Mol Cell* *1*, 991-1000.
- Neve, J., Patel, R., Wang, Z., Louey, A., and Furger, A.M. (2017). Cleavage and polyadenylation: Ending the message expands gene regulation. *RNA biology* *14*, 865-890.
- Nojima, T., Gomes, T., Carmo-Fonseca, M., and Proudfoot, N.J. (2016). Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nat Protoc* *11*, 413-428.
- Nojima, T., Tellier, M., Foxwell, J., Ribeiro de Almeida, C., Tan-Wong, S.M., Dhir, S., Dujardin, G., Dhir, A., Murphy, S., and Proudfoot, N.J. (2018). Deregulated expression of mammalian lncRNA through loss of SPT6 induces R-Loop formation, replication stress, and cellular senescence. *Mol Cell* *72*, 970-984.e977.
- Pelechano, V., and Steinmetz, L.M. (2013). Gene regulation by antisense transcription. *Nat Rev Genet* *14*, 880-893.

- Pfeffer, S., Zavolan, M., Grässer, F.A., Chien, M., Russo, J.J., Ju, J., John, B., Enright, A.J., Marks, D., Sander, C., *et al.* (2004). Identification of virus-encoded microRNAs. *Science* 304, 734-736.
- Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA Polymerase II juggernaut. *Science* 352, aad9926.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.
- Ramírez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44, W160-165.
- Ray, J., Munn, P.R., Vihervaara, A., Lewis, J.J., Ozer, A., Danko, C.G., and Lis, J.T. (2019). Chromatin conformation remains stable upon extensive transcriptional changes driven by heat shock. *Proceedings of the National Academy of Sciences* 116, 19431.
- Reimer, K.A., Mimoso, C.A., Adelman, K., and Neugebauer, K.M. (2021). Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis. *Mol Cell*.
- Richter, K., Haslbeck, M., and Buchner, J. (2010). The Heat Shock Response: Life on the Verge of Death. *Mol. Cell* 40, 253–266.
- Robbins, E., Pederson, T., and Klein, P. (1970). Comparison of mitotic phenomena and effects induced by hypertonic solutions in HeLa cells. *J Cell Biol* 44, 400-416.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.
- Rosa-Mercado, N.A., Zimmer, J.T., Apostolidi, M., Rinehart, J., Simon, M.D., and Steitz, J.A. (2021). Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression. *Mol Cell*.
- Roth, S.J., Heinz, S., and Benner, C. (2020). ARTDeco: automatic readthrough transcription detection. *BMC Bioinformatics* 21, 214.
- Rutkowski, A.J., Erhard, F., L'Hernault, A., Bonfert, T., Schilhabel, M., Crump, C., Rosenstiel, P., Efstathiou, S., Zimmer, R., Friedel, C.C., *et al.* (2015). Widespread disruption of host transcription termination in HSV-1 infection. *Nat Commun* 6, 7126.

- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148, 458-472.
- Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* 3, 1101-1108.
- Schofield, J.A., Duffy, E.E., Kiefer, L., Sullivan, M.C., and Simon, M.D. (2018). TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat Methods* 15, 221-225.
- Schwalb, B., Michel, M., Zacher, B., Fruhauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science* 352, 1225-1228.
- Shah, N., Maqbool, M.A., Yahia, Y., El Aabidine, A.Z., Esnault, C., Forné, I., Decker, T.M., Martin, D., Schüller, R., Krebs, S., *et al.* (2018). Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals. *Mol Cell* 69, 48-61.e46.
- Shalgi, R., Hurt, J.A., Lindquist, S., and Burge, C.B. (2014). Widespread inhibition of posttranscriptional splicing shapes the cellular transcriptome following heat shock. *Cell Rep* 7, 1362-1370.
- Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* 33, 365-376.
- Skaar, J.R., Ferris, A.L., Wu, X., Saraf, A., Khanna, K.K., Florens, L., Washburn, M.P., Hughes, S.H., and Pagano, M. (2015). The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Res* 25, 288-305.
- Skourti-Stathaki, K., Proudfoot, Nicholas J., and Gromak, N. (2011). Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. *Molecular Cell* 42, 794-805.
- Soucek, S., Zeng, Y., Bellur, D.L., Bergkessel, M., Morris, K.J., Deng, Q., Duong, D., Seyfried, N.T., Guthrie, C., Staley, J.P., *et al.* (2016). The evolutionarily conserved polyadenosine RNA binding protein, Nab2, cooperates with splicing machinery to regulate the fate of pre-mRNA. *Mol Cell Biol* 36, 2697-2714.

- Stadelmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., Raffel, R., Sobhian, B., Severac, D., Rialle, S., *et al.* (2014). Integrator complex regulates NELF-mediated RNA Polymerase II pause/release and processivity at coding genes. *Nat Commun* 5, 5531.
- Tarn, W.Y., and Steitz, J.A. (1994). SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes Dev* 8, 2704-2717.
- Tatomer, D.C., Elrod, N.D., Liang, D., Xiao, M.S., Jiang, J.Z., Jonathan, M., Huang, K.L., Wagner, E.J., Cherry, S., and Wilusz, J.E. (2019). The Integrator complex cleaves nascent mRNAs to attenuate transcription. *Genes Dev* 33, 1525-1538.
- Ulianov, S.V., Khrameeva, E.E., Gavrillov, A.A., Flyamer, I.M., Kos, P., Mikhaleva, E.A., Penin, A.A., Logacheva, M.D., Imakaev, M.V., Chertovich, A., *et al.* (2016). Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res* 26, 70-84.
- Vihervaara, A., Duarte, F.M., and Lis, J.T. (2018). Molecular mechanisms driving transcriptional stress responses. *Nat Rev Genet* 19, 385-397.
- Vilborg, A., Passarelli, M.C., Yario, T.A., Tycowski, K.T., and Steitz, J.A. (2015). Widespread inducible transcription downstream of human genes. *Mol Cell* 59, 449-461.
- Vilborg, A., Sabath, N., Wiesel, Y., Nathans, J., Levy-Adam, F., Yario, T.A., Steitz, J.A., and Shalgi, R. (2017). Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. *Proc Natl Acad Sci U S A* 114, E8362-E8371.
- Vilborg, A., and Steitz, J.A. (2017). Readthrough transcription: How are DoGs made and what do they do? *RNA Biol* 14, 632-636.
- Wang, X., Hennig, T., Whisnant, A.W., Erhard, F., Prusty, B.K., Friedel, C.C., Forouzmand, E., Hu, W., Erber, L., Chen, Y., *et al.* (2020). Herpes simplex virus blocks host transcription termination via the bimodal activities of ICP27. *Nat Commun* 11, 293.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (New York:Springer-Verlag).
- Wiesel, Y., Sabath, N., and Shalgi, R. (2018). DoGFinder: a software for the discovery and quantification of readthrough transcripts from RNA-seq. *BMC Genomics* 19, 597.
- Withers, J.B., Li, E.S., Vallery, T.K., Yario, T.A., and Steitz, J.A. (2018). Two herpesviral noncoding PAN RNAs are functionally homologous but do not associate with common chromatin loci. *PLoS Pathog* 14, e1007389.

- Woo, Y.H., and Li, W.-H. (2012). Evolutionary Conservation of Histone Modifications in Mammals. *Molecular Biology and Evolution* 29, 1757-1767.
- Wu, Y., Albrecht, T.R., Baillat, D., Wagner, E.J., and Tong, L. (2017). Molecular basis for the interaction between Integrator subunits IntS9 and IntS11 and its functional importance. *Proc Natl Acad Sci U S A* 114, 4394-4399.
- Wuarin, J., and Schibler, U. (1994). Physical isolation of nascent RNA chains transcribed by RNA Polymerase II: evidence for cotranscriptional splicing. *Mol Cell Biol* 14, 7219-7225.
- Xie, M., Zhang, W., Shu, M.D., Xu, A., Lenis, D.A., DiMaio, D., and Steitz, J.A. (2015). The host Integrator complex acts in transcription-independent maturation of herpesvirus microRNA 3' ends. *Genes Dev* 29, 1552-1564.
- Yao, R.-W., Wang, Y., and Chen, L.-L. (2019). Cellular functions of long noncoding RNAs. *Nature Cell Biology* 21, 542-551.
- Zhang, H., Rigo, F., and Martinson, H.G. (2015). Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. *Mol Cell* 59, 437-448.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137.
- Zimmer J.T., Rosa-Mercado N.A., Canzio D., Steitz J.A., Simon M.D. (2021). STL-seq reveals distinct roles for release and termination from the promoter-proximal pause site. *Submitted*.