NGS Applications Elucidate the Role of CDK7 and CDK9 in Global Gene Expression

by

Jessica M. Rodino

B.S., University of Nebraska at Omaha, 2020

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Master of Science

Department of Biochemistry

2023

Committee Members:

Dylan Taatjes

Jennifer Kugel

Edward Chuong

Rodino, Jessica M. (M.S., Biochemistry)

NGS Applications Elucidate the Role of CDK7 and CDK9 in Global Gene Expression

Thesis directed by Dylan J. Taatjes, PhD

Abstract

RNA polymerase II (*RNAPII)-mediated gene expression is spatially and temporally regulated by auxiliary protein factors to ensure that transcription occurs at the right place and time on the human genome. Two cyclin-dependent kinases, CDK7 and CDK9, play a major role in this regulation at all three phases of transcription: initiation, elongation, and termination. This regulation is in part achieved by CDK7 and CDK9's phosphorylation of the RNAPII C-terminal domain (CTD), at serines 2 (CDK9), 5 (CDK7), and 7 (CDK7) of the CTD's 52 repeats with the consensus heptad sequence $YS_2PTS_5PS_7$. This CTD phosphorylation "code" is thought to play a major role in recruiting protein factors to RNAPII that help regulate the different phases of transcription, effectively ensuring that the process proceeds from one checkpoint to the next. However, these kinases' role in RNAPII transcription remains poorly understood due to historically limited technologies with which to inhibit and study CDKs. In this thesis, specific, next-generation, small molecule inhibitors SY-5609 and KB-0742 were used to inhibit CDK7 and CDK9 respectively in IFN-γ stimulated HCT116 cells. ChIP-Seq experiments with antibodies for pSer2, 5, 7, and total RNAPII were then performed to determine the genomic locations of differentially phosphorylated forms of RNAPII when CDK7 and CDK9 were inhibited.*

This work showed that CDK9 inhibition results in an accumulation of RNAPII at gene 5' ends, consistent with previous understandings of this kinase's role in RNAPII promoter-proximal pause release. This work also revealed that pSer2 marks are found at the 5' ends of many human genes, a distribution which is distinct from yeast and has been identified but ignored in previous literature and about which little is known. Additionally, pSer7 is more negatively impacted than pSer5 by CDK7 inhibition. These results suggest that CDK7 and CDK9 may be subject to levels of compensation by other kinases, and that inhibition of CDK7

and CDK9 may result in phosphatase dysregulation that impacts CTD phosphorylation patterns. While preliminary, these results raise new questions about the distribution of RNAPII CTD phosphomarks on human genes and the regulatory roles of the kinases that deposit them.

ACKNOWLEDGEMENTS

I am incredibly fortunate to have found myself in the presence of great friends, family, and colleagues, whose personal and professional contributions to my success cannot be overstated.

First, thanks to my wonderful, selfless, partner, Nick Jaton. Your continued support of my goals over the last 2.5 years has been instrumental. Thank you for always pushing me to be the best version of myself, and for doing the cooking almost every night. Thank you to Apollo the dog for being good company during long writing sessions.

To my oldest friends: Ariana Haddad, Emily Young, Clare Longacre, McKensi Uecker. Thanks for sticking with me through the years (or decades) and pretending to understand my job.

I was fortunate enough to make a slew of friends in this department and in Boulder at large. Thank you for talking science with me, commiserating with me, and laughing at my bad jokes. I'm a chronically infrequent communicator, but you all mean the world to me.

I entered graduate school with an uncharacteristically large cohort of scientists who are kinder and more intelligent than I could have ever anticipated. You have all weathered the ups and downs of our unique circumstances with grace and composure. I am so excited to see what you become.

My scientific journey would never have materialized without the mentorship of Dr. David Manning and Dr. Claudia Rauter at the University of Nebraska at Omaha. Thanks for letting me get my hands dirty (literally), and providing the patience and guidance necessary to prepare me for graduate research. Thank you to Dr. John Conrad for showing me that chemistry could be fun.

I could not have asked for a better group of coworkers than those of the Taatjes Lab. Thank you to Dr. Taylor Jones and soon-to-be Dr. Olivia Luyties for your mentorship that made this work possible. Thank you to all lab members for your generosity, scientific accumen, and magnificent sense of humor. I will remember our time on the veranda fondly.

Thank you to my committee members, Dr. Jen Kugel and Dr. Edward Chuong for your careful and professional scientific assessment of the work contained herein.

Thank you to Dr. Dylan Taatjes for being my dedicated mentor through this entire process. Your persistent optimism, professionalism, and consideration makes you a good manager and a better scientist. Thank you for being on my team.

CONTENTS

TABLES

FIGURES

1. INTRODUCTION

1.1 Central Dogma and Gene Expression. The human body contains roughly 37 trillion cells and at least 200 different cell types[1–4]. Cell types vary in function and morphology, and even cells of the same type display an impressive amount of heterogeneity depending on their unique epigenetic modifications and cellular signaling[4]. Despite these differences, the majority of these cells contain the same genetic information. How, then, is such heterogeneity established when the blueprint of life remains constant? The answer lies in the process of gene expression, which allows different genes to be expressed in response to internal and external signals. This facilitates cell differentiation, cell identity, and all subsequent cellular function. Aberrant gene expression often results in cancer and developmental disorders. Clearly, this process is critical to sustaining human life.

The process of gene expression involves three main steps, known as the Central Dogma of biology[5] (Figure 1). The first is the process of replication, in which DNA is copied prior to cell division to ensure that all cells contain the appropriate genomic blueprint. The second step in this process is the conversion of DNA to RNA, which is known as transcription. Finally, much of the RNA code is used to build proteins, the functional molecules of the cell, during the process of translation. The translation machinery is in part made up of ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs), which do not code for proteins but do facilitate protein production. These are part of a larger class of non-coding RNAs, which make up a bulk of the transcriptome and have been

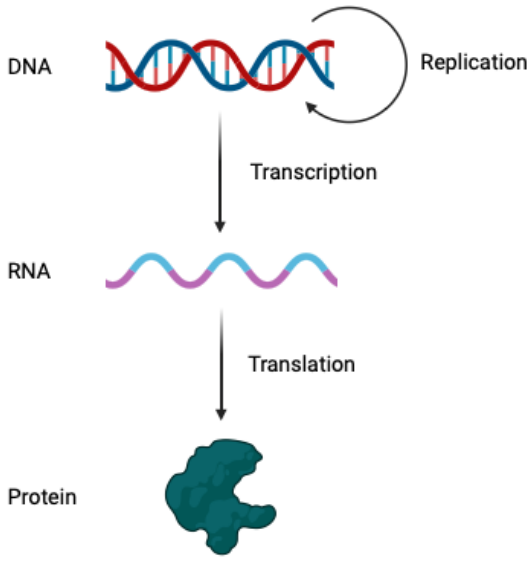


Figure 1: Central Dogma of Biology. Created in BioRender.

implicated in a multiple aspects of gene expression in recent years[6]. Without controlled gene expression programs, DNA would be converted to RNA for protein production at random. To prevent this, expression is a tightly regulated process. Much of this regulation occurs during transcription.

The simple definition of transcription is stated above: DNA is converted into RNA. Eukaryotic transcription is performed by a class of proteins called RNA polymerases. RNA Polymerase I (RNAPI) transcribes genes that produce ribosomal subunits (rRNAs), while RNAPIII transcribes genes that encode tRNAs and other small RNAs[7]. Eukaryotic transcription of protein coding genes is performed by RNA Polymerase II (RNAPII) and is the focus of this work. The RNAPII active site creates the ideal environment for the formation of RNA by matching free nucleotide triphosphates (NTP's) to their corresponding DNA base via phosphodiester bonds. However, this process must occur at the correct position on a three billion base-pair genome at precisely the right time. RNAPII must be guided to the correct DNA site, and it also cannot melt the double-stranded DNA structure to start transcription on its own. To accomplish these feats, RNAPII is accompanied by a variety of auxiliary factors that spatially and temporally regulate transcription.

1.2 Promoters and Enhancers:

Promoters and enhancers are key stretches of genomic DNA whose presence is required for regulated transcription to occur in vivo. The promoter sequence serves as a docking point for the pre-initiation complex (PIC), and the promoter sequence is located around the transcription start site, or TSS. Human promoters generally fall into one of two categories: those that contain a TATA box, and those that are GC-rich. Though the adherence of a promoter to these sequence motifs is highly variable and not required for function[8], TATA boxes tend to be present on inducible, tissue-specific genes, while GC-rich promoters are generally found on housekeeping genes or those involved in embryo patterning and morphogenesis[9,10].

Enhancer identity and function is less clearly defined than promoters. Enhancers can exist many kilobases away from the TSS and serve as cis-acting activators of a target gene[11]. Enhancers are transcribed to produce eRNAs, which are short and unstable compared to their mRNA counterparts. TFs, chromatin remodelers, and GTFs also interact with enhancers. Enhancers can be divided into two categories, traditional enhancers and super-enhancers. The latter consists of clusters of traditional enhancers that typically drive high levels of transcription[12]. As enhancers are often located a great distance away from the promoter/TSS, one model for these interactions involves a process known as enhancer-promoter looping. This looping facilitates the formation of topologically associated domains (TADs) that provide another axis of regulation to gene expression at the chromatin level that is just beginning to be understood[11,13].

1.3 PIC Components and Assembly:

The PIC (Figure 2) positions RNAPII at the TSS and is made up of general transcription factors (GTFs) and the promoter DNA to which they are bound. The GTFs include: TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, Mediator, and RNAPII. These proteins and protein complexes are involved in most, if not all, RNAPII transcription, though TFIID's composition may be promoter-specific with the exception of TATA-binding protein (TBP)[14–16]. Nonetheless, GTFs assemble to form the PIC that is required for all RNAPII-mediated transcription in cells. PIC assembly is largely sequential in nature[17–20]. First, TFIID is recruited and binds to the DNA template at the TATA and BRE sequences via its TBP

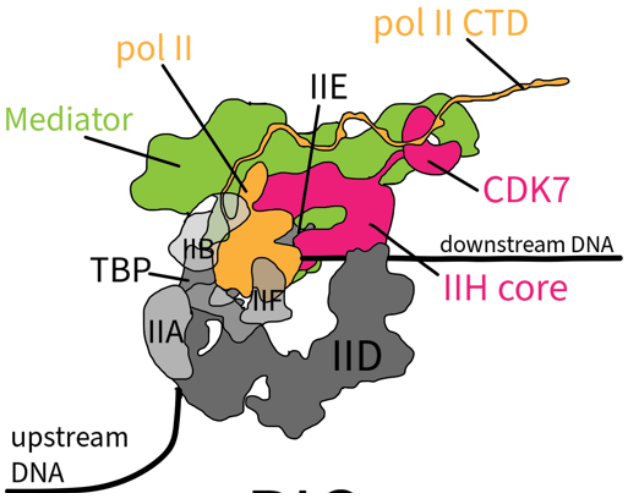


Figure 2: Schematic depicting the structure of the Pre-Initiation Complex.

subunit[21–23]. In TATA-less promoters, TFIID lobe C binds to the downstream promoter element (DPE) located at +10 nucleotides from the TSS[22]. Following this, TFIIA binds to TBP, allowing it to displace TFIID associated factors (TAFs) that inhibit PIC formation, likely as a regulatory measure[18,24]. TFIIB also binds with and stabilizes TBP and TFIIA, providing a docking point for RNAPII and TFIIF[25–27]. This event is followed by the recruitment of TFIIE and TFIIH, the former of which serves to activate the latter's CDK7 kinase activity and XPB translocase activity[28–30]. Finally, Mediator complex forms extensive contacts with RNAPII and TFIIH that allow it to stabilize the PIC and position and activate CDK7 for RNAPII CTD phosphorylation[31–34]. Mediator also works in conjunction with the Cohesin complex to regulate enhancer-promoter looping, acting as a bridge via its direct binding to enhancers[35–38]. This effectively links cell signaling events to transcription initiation.

1.4 Cyclin-Dependent Kinases:

One class of protein modulators, cyclin-dependent kinases, have captured the attention of pharmaceutical researchers as attractive and druggable anti-cancer targets due to their ubiquity in global transcription regulation. Cyclin-dependent kinases (CDK) (Figure 3) perform reversible phosphorylation on substrates that regulate growth, development, and gene expression[39–43]. These kinases are positively regulated by their associations with their cyclin partners. All CDKs have a cyclin partner, many of which overlap between kinases. The cyclin partner coordinates the T loop in the CDK active site to position it for phosphorylation, which activates the kinase[39,44].

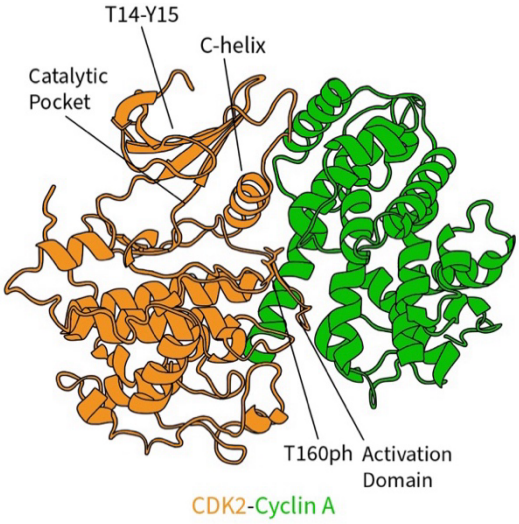


Figure 3: Cyclin-Dependent Kinase Structure. Broadly, most CDKs and cyclins share a similar structure that retains key elements important to their functionality. Figure adapted from Malumbres et al. 2014.

Broadly, there are two classes of cyclin-dependent kinases: cell-cycle CDKs and transcriptional CDKs. The cell cycle CDKs are made up of CDK 1-6 and 14-18. The transcriptional CDKs consist of CDK7-9, 12, and 13. CDK12 and 13 are homologous and appear to play a role in elongation, splicing, and 3' end processing [45–51]. CDK8 makes up part of the Mediator kinase module (MKM) which is thought to regulate stimulus response-mediated transcription through enzymatic and scaffolding mechanisms during initiation that are poorly defined[52–57]. CDK7 and CDK9, however, appear to be required for all transcription and act as regulators through the entire process from initiation to termination (Figure 4, Table 1). As such, CDK7 and CDK9 have attracted significant attention as anti-cancer drug targets due to their roles in global transcription, and are the main focus of this thesis.

1.4.1 CDK7: CDK7 is unique in that it is involved both in the cell cycle and the regulation of gene expression. It maintains such a broad functionality because it is a CDK-activating kinase (CAK), and it phosphorylates other CDKs to upregulate their activity. CDK7 is known to phosphorylate CDKs 1,2, and 4/6[58], a role that has implicated CDK7 as a viable drug target in CDK4/6 inhibitor resistant breast cancer[59,60]. CDK7 is often associated with its roles in the cell cycle, but it also plays a prominent role as a global regulator of transcription. In transcription, CDK7, Cyclin H, and MAT1, make up the CAK of TFIIH, a general transcription factor. MAT1 helps promote CAK formation[61,62] and CAK association with TFIIH[62–64]. It also stablilizes interactions between RNAPII and the PIC through its core PIC interactions[62,65]. CDK7 phosphorylates the RNAPII CTD at the Ser5 and Ser7 positions in addition to acting as a CAK for a variety of other CDKs[66,67]. These phosphorylation marks play a major role in recruiting RNA processing factors involved in capping, splicing, and termination as well as chromatin modifiers (SETD1 and SETD2) to actively transcribed regions[68,69].

1.4.2 CDK9: Of the transcriptional kinases, CDK9 is most thoroughly characterized. This is because CDK9 plays a critical role in both basal and stimulus-response mediated transcription[77–79], unlike CDK12/13 whose impacts seem to be largely relegated to the DNA damage response and snoRNA genes, respecitively[80]. Unlike CDK7, CDK9 does not possess CAK activity, making its functions and targets easier to identify. CDK9 associates with Cyclin T to make up P-TEFb, which is responsible for depositing pSer2 on the RNAPII CTD as well as phosphomarks on a slew of other substrates[81,82]. P-TEFb associates with other proteins to form elongation complexes, which appear to be important for RNAPII processivity and termination.

	Cyclin Partner	Complex	Substrates	Role in Transcription	Inhibitors[70]
CDK7	H	TFIIH	CDK9[71], RNAPII CTD	Promoter Escape, Pausing, Pause Release[71]	**SY-351**[72]**, SY-5609**[73], CT7001, LDC4297, SY-5102, YKL-5-124
CDK9	T	P-TEFb	RNAPII CTD	Pause Release, Elongation, Termination[74,75]	**KB-0742**[76], AZD4573, VIP152, Compound 30i, iCDK9, JSH-150, MC180295, NVP2, THAL-SNS-032

Table 1: Basic roles of CDK7 and CDK9 and their corresponding complexes. Inhibitors in bold were used in this research.

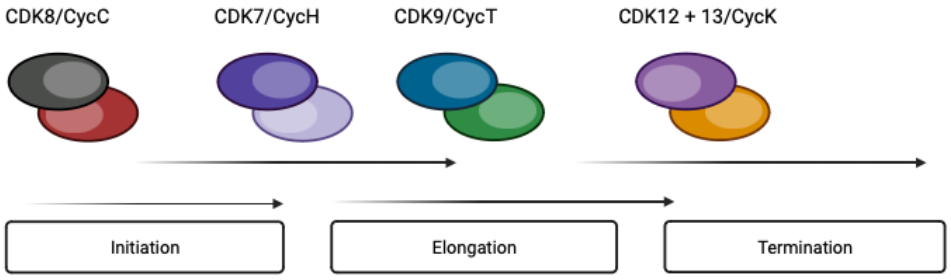


Figure 4: Kinases Regulate Each Phase of Transcription. Figure adapted from Fischer et al. 2019

CDK7 and CDK9 appear to play important regulatory roles in all phases of transcription: inititation, elongation, and termination. Because these kinases govern so many transcriptional events, a broad understanding of the major components of transcription is necessary.

1.5 CDK7 and CDK9 regulate various stages of transcription:

Once the PIC is assembled at the promoter, the process of transcription initiation can occur when NTPs are present. Two key regulatory checkpoints precede productive elongation: promoter escape and promoter-proximal pausing. These steps allow for tight control of gene expression programs through CDK7 and CDK9 activity.

1.5.1 Promoter Escape: Precise PIC assembly is important to correctly position RNAPII for transcription, and PIC contacts with polymerase are therefore extensive[18,83]. This creates an energetic barrier that must be overcome for transcription to proceed[84]. TFIIH's 5' to 3' translocase activity plays an important role in this process by generating torsional strain in the nontemplate strand, which facilitates opening of the DNA duplex and feeds the the template into the RNAPII active site to establish the open complex[85–87]. After RNAPII has transcribed 8-12 nucleotides, TFIIB dissociates from the PIC to open the RNAPII RNA exit channel, and a DNA:RNA hybrid forms in the RNAPII active site akin to that of the elongation complex[18,88]. Furthermore, the RNAPII CTD makes extensive contacts with Mediator within the PIC[18,83]. Phosphorylation of the CTD by CDK7 during

initiation may disrupt these contacts and help to free RNAPII from the PIC[31,89,90]. Mediator also plays a role in promoter escape through its contacts with sequence-specific transcription factors. These contacts likely restructure Mediator-RNAPII contacts to make them permissive to promoter escape[18,33,91]. The CDK8 module is also hypothesized to play a role in this process, as its binding to Mediator is mutually exclusive with that of RNAPII[52–56]. Though the Taatjes lab possesses preliminary data that portends this role for the CDK8 module, this has never been confirmed[92]. Because of this, current research implicates CDK7 as the primary kinase involved in promoter escape.

1.5.2 Promoter-proximal pausing: Following promoter escape, RNAPII typically pauses after transcribing 20-80 nucleotides. This is the promoter-proximal pause, and it is understood to be the rate-limiting step in transcription[93–95]. The pause is thought to serve several functions. It likely helps establish a permissive (that is, open) chromatin architecture for transcription by preventing a nucleosome from occupying the promoter-proximal region[96]. The promoter-proximal paused RNAPII also allows for rapid transcriptional activation in response to signaling events[97]. In the case of the latter, pausing has been observed at rapidly induced genes[98,99], but appears to be more enriched at constitutively expressed genes that produce components of signaling pathways[100]. This suggests that pausing dually regulates signal-response through both signal transduction pathways and transcription itself. Pausing also likely provides an axis for the coordination of mRNA processing factors such as the 5' capping enzyme. This enzyme interacts with pSer5 of the RNAPII CTD[101] and caps mRNA when it is around 20-30 nucelotides in length[102,103], concomitant with the pause window. It is therefore hypothesized that pausing provides a window of opportunity for capping machinery to associate with RNAPII and the nascent mRNA.

To facilitate pausing, TFIIE and TFIIF dissociate from RNAPII after it escapes the promoter. This likely occurs to allow two transcriptional cofactors, NELF and DSIF, to bind to polymerase in their stead[104]. NELF and DSIF stabilize pausing[104,105] and help regulate the

proper formation of early elongation complexes through contacts with 5' capping machinery[106,107]. DSIF binds to polymerase first, and appears to play a role in the recruitment of NELF[108]. While NELF association with polymerase is only associated with establishing the pause, DSIF can act as a positive or a negative elongation factor. This dual functionality is conferred by the phosphorylation of DSIF's Spt5 subunit by CDK9 during pause release[109,110]. When DSIF is phosphorylated, it becomes a positive elongation factor[111,112]. CDK9 also phosphorylates NELF, and this combined with Spt5 phosphorylation results in NELF dissociation from RNAPII[110,113], allowing polymerase to proceed into productive elongation.

Important questions about promoter proximal pausing remain unanswered. Research from the Taatjes lab showed that NELF and DSIF were not required to reconstitute pausing in vitro. Instead, the TFIID TAF1/2 subunits were implicated as required pause regulators during both in vitro biochemical reconstitution and in vivo rapid depletion in human and Drosophila cells[114]. Work from the Shilatifard lab using an auxin-inducible degron tagged NELF further observed that NELF depletion resulted in global pause dysregulation under basal conditions, but not during the heat shock response[106]. This is in accordance with similar observations in Drosophila[115], suggesting that DSIF/NELF-independent pausing may depend on transcriptional context.

1.5.3 CDK7, CDK9, and Pausing: Clearly, CDK7 and CDK9 are important for pause regulation, but their unique functions in this process are still in many ways unknown. CDK7 likely plays a dual role in both pause establishment and pause release, though the former is less clearly understood. Chemical genetics experiments involving analog-sensitive

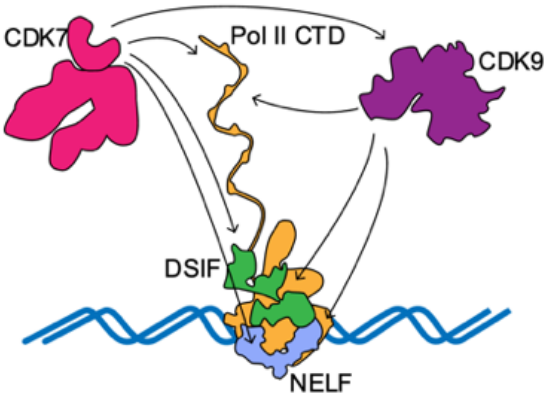


Figure 5: Phosphorylation Targets of CDK7 and CDK9 are Vital to Pause Regulation.

CDK7 (CDK7as) cell lines and kinase inhibitor THZ1 showed that inhibition of CDK7 resulted in a decrease in recruitment of DSIF and NELF as well as an attenuation of pausing[68,71,116]. Paradoxically, work from the Taatjes lab using both a CDK7as cell line and inhibition of WT CDK7 by THZ1 showed an increase in paused polymerase[69], which is perhaps reflective of its role as a CAK for CDK9. DSIF, NELF, and CDK9 have all been identified as CDK7 phosphorylation targets[71,117]. CDK9 can also phosphorylate NELF and DSIF[107,110,113], inferring that CDK7 is the master regulator of pause release in addition to pause establishment. CDK7's nuanced role in this process has confounded researchers, and it is not clear how CDK7 temporally regulates the processes of pausing and pause release. In this way, the unique contributions of both CDK7 and CDK9 to pausing as a whole are poorly understood.

1.5.4 Elongation and Termination:

Following pause release, polymerase proceeds through elongation to the 3' end of the gene.The most prominent kinase involved in elongation is CDK9, which is responsible for regulating RNAPII through elongation and termination. While CDK9 inhibition does not appear to affect RNAPII's ability to transcribe through large distances within the gene body, loss of CDK9 function does appear to cause errors in termination[74,77,118]. CDK7's roles in elongation and termination are also of great interest to researchers, suggesting that both of these kinases are important in stages of transcription beyond pausing.

1.5.5 CDK7 in Elongation and Termination: CDK7 is most frequently implicated in promoter-proximal events in transcription, but recent work suggests roles that may extend into elongation and termination. CDK7 loss of function results in delayed termination, negatively impacts 3' end formation, and changes chromatin modification patterns[69]. Errors involved with termination and 3' end formation are likely a result of CDK7's role in activating CDK9 and possibly CDK12 and CDK13[117]. Comparisons of the impact of CDK7 inhibition and CDK9 inhibition show that CDK7 impacts 3' end formation to a lesser extent than CDK9, suggesting that CDK7 function is likely indirect[68,71]. Changes in

chromatin modifications are similarly unclear. Sources have implicated CDK7 in cotranscriptional histone modification pathways, including H2Bub1, H3K4me3, and H3K36me6[68,69,71]. Work from the Taatjes lab using an analog-sensitive CDK7 cell line provided a mechanistic insight into these assertions through mass spectrometry, which revealed that the TFIIH-phosphorylated CTD is selectively bound by H3K4 methyltransferases SET1A/B and stimulates protein activity toward nucleosomes to induce H3K4me3 spreading[69]. H3K36 methyltransferase activity by SET2 was shown to be directly linked to CDK9[69], suggesting that CDK7's effect on this modification is a result of its role as a CAK[69].

RNA-Seq data produced by the Taatjes lab also suggests that chemical inhibition of CDK7 results in widespread splicing defects in the form of both abnormally reduced and increased exon skipping and increased intron retention in 11,348 instances[117]. This same study also utilized quantitative phosphoproteomics to implicate several splicing factors as CDK7 targets, most notably SF3B1. SF3B1 helps hybridize the pre-mRNA to spliceosome machinery. Furthermore, splicing changes induced by CDK7 inhibition mirrored those produced by an SF3B1 inhibitor in 1152 instances[117,119], suggesting that CDK7 acts as a regulator of splicing, in part, through SF3B1. Previous data from the field also identified CDK7 as a required component for the expression of heat shock genes in Drosophila[120]. Heat shock is known to cause splicing defects that can be rescued by heat shock proteins[121], portending a unique role for CDK7 in this process. In this way, CDK7 appears to play both direct and indirect roles in all phases of RNAPII transcription.

1.5.6 CDK9 in Elongation and Termination: As previously stated, CDK9's most prominent impacts following pause release appear to occur during termination, after RNAPII has transcribed past the polyA signal. That being said, it does aid in elongation. Because CDK9 is not a part of the PIC, it must first be precisely delivered to genomic loci. This delivery is accomplished through its associations with a variety of protein cofactors, which assemble with P-TEFb to form elongation complexes. Active P-TEFb is delivered to

promoters as apart of either the BEC (BRD4 containing Elongation Complex)[75,122] or SEC (Super Elongation Complex). While the BEC has been implicated at enhancer loci, the SEC appears to function at both promoters and enhancers[75]. Some studies have also suggested that these complexes operate cooperatively to recruit CDK9 and promote gene expression[123]. That being said, data from the Taatjes lab shows that P-TEFb binds the RNAPII CTD regardless of its phosphorylation status[69], which suggests that the presence of RNAPII is sufficient to recruit CDK9 to actively transcribed regions.

Once CDK9 is appropriately deposited at genomic loci, it can perform its roles in elongation and termination. CDK9's phosphorylation of DSIF's Spt5 subunit enhances processivity[74,124] and aids in the recruitment of elongation factors like PAF1C. PAF1C plays a critical role in elongation[125], likely by preventing NELF reassociation and aberrant RNAPII pausing[74,104]. CDK9 also simultaneously renders phosphatases inactive to maintain Spt5 phosphorylation[74]. In this way, CDK9 helps mediate the recruitment of positive elongation factors that help RNAPII maintain processivity as it transcribes through the gene body. While CDK9 is likely not absolutely required for the recruitment of these factors, its presence helps increase the efficiency of elongation.

CDK9 also directly and indirectly regulates termination. Once RNAPII transcribes past the polyA site, CDK9 activity decreases, and protein phosphatase 1 dephosphorylates Spt5[126]. This causes RNAPII to slow down. When CDK9 is inhibited, premature termination is observed, likely because Spt5 is dephosphorylated prior to the polyA site[118]. Additionally, CDK9 promotes the recruitment of Xrn2, which is a 5' to 3' exonuclease that plays a major role in the "torpedo" model of transcription termination[116,127]. Xrn2 appears to be dependent on CDK9 phosphorylation for chromatin localization[116], at which point it degrades the RNA transcribed after RNAPII passes the polyA site and the nascent RNA is cleaved. Overall, CDK9 function is important for regulating the transition of RNAPII from elongation to termination[74].

1.6 CDK7, CDK9, and Enhancers

Both CDK7 and CDK9 play a role in the regulation of enhancer transcription. This is part of what has made these kinases attractive drug targets for cancer therapeutics. Many tumors contain specific super-enhancers, which recruit TF's and GTF's to drive large amounts of gene expression related to cell identity and survival[12,68]. CDK7 is often enriched at these super-enhancers, and many CDK7 inhibitors take advantage of this to induce cytotoxic effects in tumors that are less harmful to healthy cells[68,128,129]. Enhancer transcription also likely utilizes DSIF and NELF mediated RNAPII pausing as a regulatory step, portending a role for both CDK7 and CDK9 in the process analogous to their roles at promoters[75,130]. CDK9 is present at both traditional enhancers and super-enhancers, and much like CDK7, inhibition of CDK9 with flavopiridol reduces eRNA transcription[130]. Enhancer transcription appears to be a requirement for the expression of enhancer-linked genes, thereby creating yet another axis of transcriptional regulation at which CDK7 and CDK9 mediate gene expression (Figure 6).

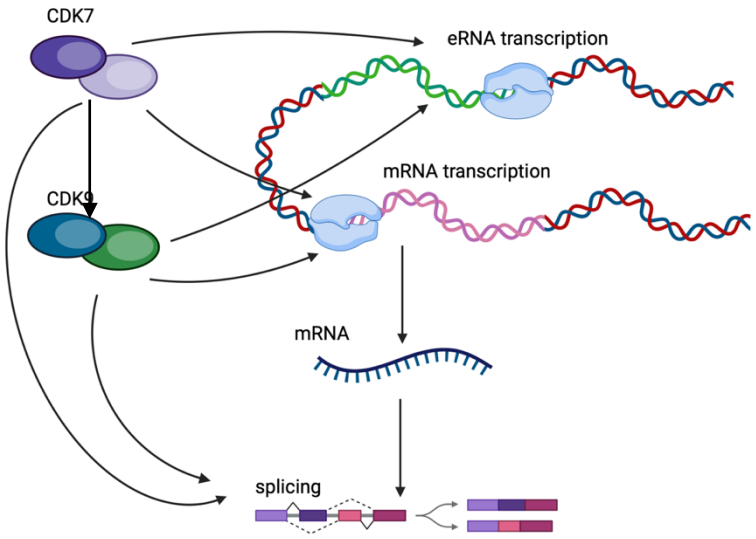


Figure 6: CDK7 and CDK9 regulate multiple axes of gene expression. In addition to mRNA transcription, CDK7 and 9 regulate enhancer transcription, which is positively correlated with expression of corresponding genes. Both kinases also module splicing events through interactions with and recruitment of splicing machinery. Taken together, these kinases have highly multifaceted roles in transcription.

1.7 RNAPII CTD Phosphorylation by CDK7 and CDK9: A Major Axis of Regulation

Transcriptional kinases target the RNAPII CTD during promoter escape, promoter proximal pausing, and elongation to modulate gene expression. The RNAPII CTD is an intrinsically disordered structure that consists of 52 amino acid heptad repeats (YSPTSPS). It also serves as a site of post-translational modifications that are hypothesized to help aid in the recruitment of various transcription cofactors, thereby regulating the procession of transcription. During initiation, the CTD is hypophosphorylated. This allows the CTD to make extensive contacts with Mediator within the PIC[131]. These contacts help position and stabilize RNAPII to prime it for transcription[132]. During promoter escape, the CTD is phosphorylated by CDK7 at the serine 5 and serine 7 positions. This is thought to disrupt contacts between Mediator and the CTD, allowing RNAPII to separate from the PIC (i.e. promoter escape) and start transcribing. Additionally, pSer5 marks help with the recruitment and positioning of 5' capping factors that stabilize the nascent transcript[133–135]. pSer2 is linked to CDK9 activity, and is thought to help release the pause and recruit factors that promote elongation and termination[74,77,136]. In yeast, pSer2 marks are largely relegated at gene 3' ends. In humans, however, evidence shows pSer2 at gene 5' ends as well as 3' ends[137–139], though this phenomenon is rarely discussed and little is known about the pSer2 function in this regard. pSer7 is also phosphorylated by CDK7 near the promoter, and is correlated with an increase in P-TEFb activity on the CTD[140]. It has

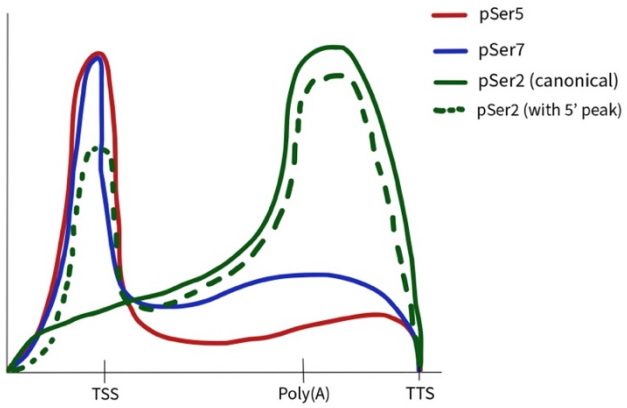


Figure 7: Canonical Phosphorylation Trends on the RNAPII CTD Throughout the Gene Body. Figure adapted from Zabrowska et al 2016. pSer2 has also been observed at the 5' ends of genes in human cells, the distribution modeled here is based off of observations made in Vihervaara et al. 2023

therefore been postulated that pSer7's role primarily involves promoting elongation and suppressing cryptic transcription[141]. Overall, the popular understanding of RNAPII CTD phosphorylation is that pSer5 and pSer7 phosphorylation is most abundant around the TSS and the pause site, while pSer2 marks steadily increase after pause release and become most prominent towards the 3' end of the gene (Figure 7). A significant body of evidence, however, indicates that the mammalian pattern of pSer2 deposition does not follow this model.

Clearly, deposition of phosphomarks by CDK7 and CDK9 is linked to these kinases' abilities to regulate the procession of transcription. CTD phosphorylation is heavily correlated with the transition of transcription between phases that correspond with the recruitment of different cofactors, indicating that it plays a major role in the control of gene expression. Understanding how CDK7 and CDK9 contribute to CTD post-translational modifications is therefore vital to ascertain how these kinases regulate transcription.

1.8 Open Questions and Significance:

CDK7 and CDK9 are key regulators of transcription during initiation, elongation, and termination. They help maintain appropriate chromatin architecture and assist in the formation of TADs that facilitate gene expression. That being said, important questions about how these kinases are regulating global gene expression remain unanswered. This is largely due to the fact that much of our understanding surrounding these kinases comes from experiments that utilized nonspecific chemical inhibitors. Because all kinases perform the same enzymatic functions, their active sites are extremely similar and developing specific chemical inhibitors for each kinase is challenging[70]. Inhibitors are typically competitive in nature and therefore must also compete with high intracellular concentrations of ATP for the kinase active site[70].

Early CDK7 inhibition studies utilized timepoints that ranged from hours to days, while pausing occurs within minutes. This provides an ample amount of time for compensatory mechanisms to return gene expression to homeostasis. As a result, CDK7's

importance in transcription, namely in pausing, is often underestimated. CDK9 experimentation faces similar shortcomings . Many inhibitor experiments that explore CDK9 function utilized DRB (5,6-dichloro-1-β-D-ribofuranosylbenzimidazole) or flavopiridol. Both DRB and flavopiridol inhibit both CDK9 and CDK7[142,143], along with other kinases. This means that much of the available data about CDK9 cannot be separated from the off-target effects of these inhibitors.

Research into CDK7's impact on splicing and 3' end processing performed in the Taatjes lab used an extremely selective, covalent CDK7 inhibitor called SY-351, and as such is not limited by the same nonspecificity constraints often seen in other kinase experiments[72,117]. However, these RNA-seq experiments utilized short reads that were typically 150-300 bp in length. While informative, short read sequencing provides only a limited view of abnormal splicing isoforms that can occur when CDK7 function is impaired. Reads shorter than 300 bp are too short to detect 70% of structural variation in the human genome that impacts sequences longer than 50 bp, with sequences 2kb in length or less being especially impacted[144,145]. Furthermore, 15% of the human genome is inconducive to short-read sequencing due to repetitive genome elements or abnormal GC content[146]. Because of these limitations, our understanding of CDK7's impact on splicing is incomplete.

Clearly, key questions remain. How exactly do CDK7 and CDK9 regulate pausing and gene expression? Do they possess redundant or unique functionality? How does this influence global transcription; are inducible genes that respond to signaling pathways regulated differently than those that are constitutively expressed? How do CDK7/9 globally regulate transcription through phosphorylation patterns on the RNAPII CTD?

The work described in the subsequent chapters seeks to remedy the gaps in our knowledge about how CDK7 and CDK9 regulate global gene expression. First, I use potent and specific "next generation" inhibitors to link CDK7 and CDK9 activity to gene expression through the lens of RNAPII phosphorylation via ChIP-seq. I also perform long-read RNA-seq with cells treated with a CDK7 inhibitor to assess this kinase's role in splicing on a genome-

wide level, at a resolution never before seen in the field (Figure S34-S38). Taken together, these experiments reveal key aspects of CDK function that inform how we understand global gene expression.

2. METHODS

2.1 ChIP-seq:

Table 2: ChIP-seq Reagents/Instruments:

Reagent	Supplier	Cat. #
McCoy's 5A Medium	Sigma-Aldrich	M4892
Fetal Bovine Serum	PEAK Serum	PS-FB3
Penicillin/Streptomycin	Invitrogen	15140
16% Formaldehyde, Methanol Free, Ultra Pure	Polysciences Inc.	18814-10
SY-351	Syros Pharmaceuticals	Proprietary
SY-5609	Syros Pharmaceuticals	Proprietary
KB-0742	TargetMol	T9446
IFNγ	Gibco	PHC4031
truChIP Chromatin Shearing Kit with Formaldehyde	Covaris	520154
M220 Focused Ultrasonicator	Covaris	500295
Millitube 1 mL AFA Fiber	Covaris	520130
Protein G Dynabeads	Invitrogen	10003D
Proteinase K	NEB	P8107S
Rpb1 NTD D8L4Y Rabbit mAb	Cell Signaling Technologies	14958S
Anti-RNA Polymerase II subunit B1 (phosphor-CTD Ser-2) Antibody, clone 3E10	Millipore-Sigma	04-1571-I
Anti-RNA Polymerase II subunit B1 (phosphor-CTD Ser-5) Antibody, clone 3E8	Millipore-Sigma	04-1572-I

Anti-RNA Polymerase II subunit B1 (phosphor-CTD Ser-7) Antibody, clone 4E12	Millipore-Sigma	4-1570-I
5PRIME Phase Lock Gel Light 2 mL Tubes	Quantabio	2302820
KAPA Hyperprep Kit (KK8502)	Roche	07962347001
KAPA Unique Dual Indexed Adapter Library (KK8726)	Roche	08861862001
KAPA Pure Beads (KS8001)	Roche	08218030001
D1000 ScreenTape	Agilent	5067-5582
D1000 Sample Buffer	Agilent	5190-6502
D1000 DNA Ladder	Agilent	5067-5586
2200 Tapestation	Agilent	G2964AA
PowerUp SYBR Green Master Mix	Applied Biosystems	A25742

Table 3: ChIP-Seq Buffers and Protease Inhibitors:

NRO Buffer:

Substance	**Concentration**
Tris HCl pH 8	10 mM
NaCl	10 mM
$MgCl_2$	4 mM
NP-40	0.5%
Protease Inhibitors	1x

Block Solution:

Substance	**Concentration**
PBS	1x
BSA	0.5%

Covaris Lysis Buffer B (used to determine final concentration of SDS in IP):

Substance	**Concentration**
Tris HCl pH 7.6	10 mM
EDTA	1 mM
SDS	0.1%
Protease Inhbitors	1x

IP Buffer:

Substance	**Concentration**
Tris HCl pH 8	15 mM
NaCl	150 mM
EDTA	1 mM
Triton X-100	1%
Protease Inhibitors	1x

IP Concentrate:

Substance	**Concentration**
Tris HCl pH 8	33.75 mM
NaCl	750 mM
EDTA	1 mM
Triton X-100	5%
Protease Inhibitors	1x

RIPA Wash Buffer:

Substance	**Concentration**
Tris HCl pH 8	20 mM
NaCl	500 mM
EDTA	1 mM
Triton X-100	1%

SDS	0.1%
Protease Inhibitors	1x

LiCl Wash Buffer:

Substance	**Concentration**
Tris HCl pH 8	20 mM
LiCl	500 mM
EDTA	1 mM
Sodium deoxycholate	1%
NP-40	1%
Protease Inhibitors	1x

TE Salt Buffer:

Substance	**Concentration**
Tris HCl pH 8	10 mM
EDTA	1 mM
NaCl	50 mM

Proteinase K Buffer:

Substance	**Concentration**
Tris HCl pH 8	10 mM
EDTA	1 mM
SDS	1%

Protease Inhibitors*:

Substance	**Concentration**
Aprotinin	3-7 Trypsin Inhibitor Units (TIU)/mg protein
Sodium Metabisulfite	1 M
Benzamidine	1 M
PMSF	0.25 M

*Solutions treated with protease inhibitors were treated with a final concentration of 1 mM Sodium Metabisulfite, 1 mM Benzamidine, 0.25 mM PMSF, and 0.003-0.007 TIU's/mg aprotinin.

Cell Culture:

HCT116 cells (ATCC# CCL-247) were grown to 70-90% confluency in McCoy's 5A medium and supplemented with 10% Fetal Bovine Serum and 1% Penicillin/Streptomycin antibiotic at 37°C and 5% CO_2.

Table 4: Inhibitor and Interferon Concentrations:

Substance	**Stock Concentration**	**Final Concentration In Cells**
SY-5609	10 mM	50 nM
KB-0742	10 mM	100 nM
IFNγ	1 mg/mL	10 ng/μL

Table 5: ChIP-Seq Treatment Scheme:

	DMSO	SY-5609	KB-0742	IFNγ
Treatment Group 1	X			X
Treatment Group 2		X		X
Treatment Group 3			X	X
Treatment Group 4		X	X	X

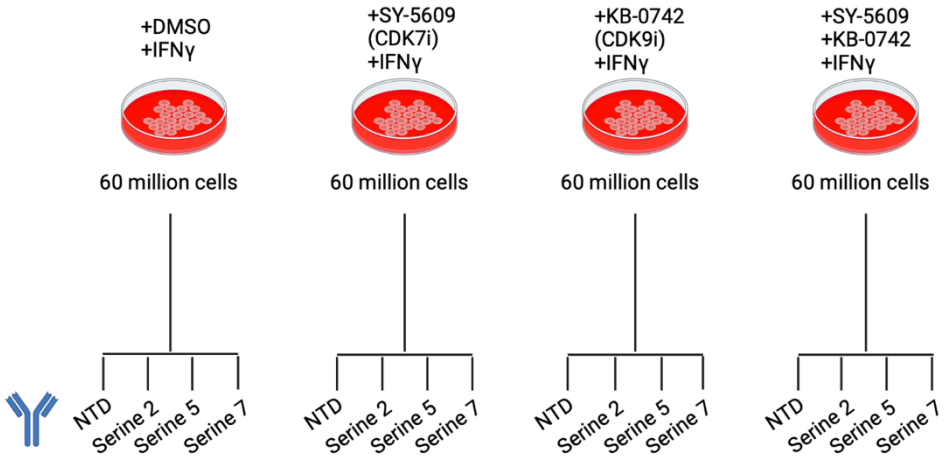


Figure 8: ChIP-Seq experimental conditions. Cells were treated with DMSO, SY-5609, KB-0742, or SY-5609 and KB-0742 (Groups 1-4) to produce two biological replicates of ~60 million cells/condition. Each treatment group was subdivided into four IPs. All cells were treated with IFNγ. Created in Biorender.

Cells treated with SY-5609 received a final concentration of 50 nM. Cells treated with KB-0742 received a final concentration of 100 nM. Working concentrations of all inhibitors were created to ensure an equal volume of drug was applied to all conditions except those treated with both inhibitors. Control cells received an equivalent volume of DMSO. Cells were incubated with either inhibitor or DMSO for 30 minutes at 37°C and 5% CO_2, at which point IFNγ was added to the same media to a final concentration of 10 ng/µL. Cells were incubated with interferon for 45 minutes at 37°C and 5% CO_2.

Crosslinking and Nuclei Isolation

Cells were crosslinked with a final concentration of 1% methanol-free formaldehyde for ten minutes while shaking at 145 RPM on an orbital shaker. Cells were quenched with 2.5 M Glycine to a final concentration of 125 mM for 8-10 minutes. Cells were harvested, washed with ice cold PBS (no calcium, no magnesium), and incubated in NRO buffer at a concentration of 80 µL/1 million cells for five minutes twice. Nuclei were pelleted, frozen in liquid nitrogen, and stored at -76°C for future use.

IP Prep

50 µL of Protein G Dynabeads/reaction was aliquoted into two tubes. Beads were washed with 1 mL of Blocking Buffer three times, then resuspended in 250 µL of blocking buffer/reaction. 6 µg/reaction of desired antibody was allowed to bind to the beads for 10 mins at room temperature with nutation, followed by nutation at 4°C for 3-6 hours.

Chromatin Shearing:

Nuclei were thawed on ice for 1.5-3 hours. Once pellets thawed, nuclei were resuspended in 1 mL Covaris Lysis Buffer B + 1x Protease Inhibitors and incubated for 10 minutes at 4°C with nutation. Nuclei were then washed with 1 mL Covaris Wash Buffer C + 1x Protease Inhibitors and incubated for 10 minutes at 4°C with nutation. Nuclei were washed with 500 µL Shearing Buffer D3 + 1x Protease Inhibitors and resuspended in 1 mL of the same solution per ~20 million cells. Nuclei were then aliquoted into AFA millitubes (1 mL/tube) and sheared on the Covaris M220 Focused Ultrasonicator for 12 minutes at 6°C, duty factor 10%, peak power 75, 200 cycles/burst. Sheared chromatin was divided into 750-780 µL aliquots (~15 million cells/aliquot) into pre-chilled Eppendorf tubes, frozen in liquid nitrogen, and stored at -76°C for future use if not immediately proceeding to the IP step.

IP Conditions:

If needed, sheared chromatin was thawed on ice for 2-3 hours. Antibody-bound beads were resuspended in 500 µL cold blocking buffer/reaction, followed by resuspension in the same volume of IP buffer/reaction. Beads were then resuspended in 50 µL IP buffer/reaction, and each reaction (50 µL beads) was aliquoted into a single 2 mL tube. 275 µL IP buffer was added to each tube, followed by 500 µL IP concentrate, and ~360 µL of sample. This means that each aliquot of approximately 720 µL of sheared chromatin from roughly 15 million cells was split into two tubes at the IP step. This volume is slightly less than that which was aliquoted during the shearing step to account for input volume and loss of volume during pipetting steps. Splitting each IP into two tubes was necessary to

dilute the amount of SDS present in the IP mixture to a final concentration of roughly 0.03% to maximize antibody-protein binding. IPs were nutated at 4°C for 12-18 hours. 10 µL of sheared chromatin from each treatment condition per IP was aliquoted into a separate tube and stored at -20°C to provide an input sample.

Washing and Reversing Crosslinks:

Beads were washed with IP Buffer, RIPA Buffer, LiCl Buffer, and TE Salt Buffer two times each to remove non-specific protein binding interactions. All washes were performed on ice and incubated for 50- 80 s. Following the final wash, beads were resuspended in 200 µL of Proteinase K Buffer containing 5 µL of 20 mg/mL Proteinase K. Beads were incubated on a Boekel Groovin' Tubes Thermomixer set to 50°C mixing at 750 RPM for one hour, or on a 50°C heat block and manually resuspended every 20 minutes. Beads were then incubated at 65°C for one hour. 190-195 µL of supernatant was removed from the beads and transferred to 0.2 mL tubes. Input chromatin was thawed and brought up to 200 µL by adding 190 µL of aforementioned Proteinase K mix. All inputs and IP's were incubated at 65°C for 17.5 hours.

DNA Isolation and Precipitation:

Samples were vortexed with an equal volume of phenol:chloroform:isoamyl (25:24:1) to generate aqueous and organic phases. Both phases were added to a phase lock tube and nutated for ten minutes at room temperature. Phases were separated via centrifugation (Phase Lock tubes contain a gel that forms a barrier between phases) and an equal volume of either phenol:chloroform:isoamyl or pure chloroform was added to the aqueous phase to perform a second DNA extraction. Phase lock tubes were again nutated for ten minutes at room temperature and centrifuged to isolate the aqueous phase. The aqueous phase was moved to a fresh tube, to which 20 µg GlycoBlue, 1/10 sample volume of 3 M NaOAc pH 5.2, 10 mM $MgCl_2$, and 2.5x total volume of 100% ethanol was added. Tubes were mixed via vortexing and incubated at -20°C for one to two days. DNA pellets were isolated via a 30 minute centrifugation at 12,000 x g. Supernatant was removed, and

pellets were washed with 750 µL of 75% ethanol. Ethanol was removed and pellets were air-dried for 5-12 minutes. Once dry, pellets were suspended in 20-30 µL of 0.1x TE Buffer. DNA concentrations were quantified via Qubit and samples were stored at -20°C.

qPCR:

ChIP efficacy was tested via qPCR. 0.4 or 0.5 ng of NTD DNA, 0.3 ng of pSer2-enriched samples, and 0.1 ng of pSer5 and pSer7 enriched sample was used per qPCR reaction. Primers (Table 6) were diluted to 400 nM per reaction. NTD, pSer5, and pSer7 samples were amplified using promoter region primers, while pSer2 samples were amplified using gene body primers. Standard curves were generated from input samples to calculate starting quantities and PCR efficiency. Signal was quantified by dividing the mean starting quantity of a ChIP sample by that of its corresponding input (Figure S3-S9).

Table 6: qPCR Primers

Gene	Region	Forward Primer (5'-3')	Reverse Primer (5'-3')
IRF1	Promoter	CGAAATGACGGCACGCAG	CCACCGAGCAATCCAAACACT
IRF1	Gene Body	GAGGAGGTGAAAGACCAGAGCA	TAGCATCTCGGCTGGACTTCGA
β2M	Promoter	GTCGCTGGCTTGGAGACAGGTG	AGCGCGACGCCTCCACTTATATT
β2M	Gene Body	GCTCTAGGAGGGCTGGCAACT	AGCTTTGAGTGCAAGAGATTGAAGA

Library Preparation:

Libraries were prepared using the KAPA Hyperprep Kit. 1 ng of total DNA from each sample was diluted in 50 µL 0.1x TE buffer and subjected to end repair and A-tailing. Adapter ligation was performed using an adapter stock concentration of 300 nM per kit instructions and was allowed to proceed for one hour at 20°C. Following ligation, a 0.8x bead cleanup was performed, and the libraries were amplified according to kit instructions. The first biological replicate of SY-5609 treated cells exhibited over-amplification on tapestation traces and was subject to 15 amplification cycles (Figure S14-S17). The second biological replicate of SY-5609 treated samples was amplified for 14 cycles, and much of

the overamplification was eliminated (Figure S18). Following amplification, libraries were subjected to two consecutive 1x bead based-cleanups to remove protein contaminants and adapter dimers. Libraries were quantified via Qubit and imaged using capillary electrophoresis on an Agilent 2200 Tapestation to identify average size. Libraries that exhibited overamplification were subjected to a reconditioning PCR, in which 16-19 µL of library was mixed with a proportional volume of primers and PCR mastermix and subjected to one cycle of amplification and a subsequent 1x bead based cleanup. This served to resolve daisy-chained DNA fragments, and reconditioned libraries exhibited the anticipated size peaks on the tapestation.

Sequencing:

Libraries were pooled by mass and concentration according to a 5:3:1 ratio of total RNAPII (NTD): phosphomarks:input samples (Figure S19-S22). Library pools were subsequently diluted to 10 nM in 30 µL according to sequencing core instructions. SY-5609 biological replicates were sequenced to a total requested depth of 600 million paired-end reads. All samples were sequenced on a NovaSeq 6000 by the Gates Institute Genomics Core at CU Anschutx.

Data Analysis:

Table 7: Software for ChIP-Seq Data Analysis

Software	Version
Nextflow[147]	19.10.0
preseq	2.0.3
bedtools[148]	2.25.0, 2.31.0
samtools[149]	1.16.1, 1.18.0
hisat2[150]	2.1.0

bbmap[151]	38.05
igvtools[152]	2.3.75
fastqc	0.11.8
rseqc[153]	3.0.0
Python[154]	3.6.3
MultiQC[155]	1.15
DeepTools[156]	3.5.2
cutadapt[157]	4.4
sambamba	1.0
UCSC-bigbedtobed	v377
conda	23.7.4

Initial sample processing and quality control was performed using the Dowell Lab ChIP-Flow pipeline (https://github.com/Dowell-Lab/ChIP-Flow). If needed, corrupted raw data files were amended using the bbmap repair function[151]. Briefly, read quality was assessed using FastQC, and adapters were trimmed from reads via bbmap. Reads were mapped to the GRCh38 reference genome (UCSC Assembly, December 2013) using hisat2. Library complexity was estimated using preseq and compressed into .tdf files for genome browser visualization using IGVTools. FastQC, preseq, and rseqc metrics were compiled into a single ouput via MultiQC. Duplicate, multimapping, and non-mapping reads were removed via sambamba. Indices for bams containing only uniquely mapped reads were generated using samtools 1.18.0. Data was then normalized to 1x read coverage via the DeepTools bamCoverage package. hg38 blacklist regions[158] were excluded from analysis. The resulting .bigwig files were used to create a matrix using DeepTools computeMatrix to calculate scores for each sample at desired genomic regions. These matrices were plotted using DeepTools plotProfile packages; bin size = 50 bp. Replicate concordance was

assessed by calculating the spearman correlation coefficient r between .bam files using the plotCorrelation function. Genomic traces from normalized bigwig files were visualized in IGV using the RefSeq MANE hg38 genome annotation[159].

Several data sets were used to generate metagene plots. The first contains IFN-γ responsive genes (n = 138) and was generated by combining the GSEA analysis[160] and genes shown to be upregulated by interferon stimulus in HCT116 cells by Steinparzer et al 2019[161] (Table S31). The second dataset contains genes typically upregulated in HCT116 cells compared to other cell lines (n = 263) and was generated from the Broad Institute's Cancer Cell Line Encyclopedia[162] and sourced from the Harmonisome 3.0 project[163] (Table S30). Finally, I separated the RefSeq MANE hg38 reference genome by size to create two distinct subsets of genes: those less than 30 kb in length, and those greater than 90 kb in length (Figure S32-S33). I did this in part through using UCSC-bigbedtobed v377.

3. RESULTS & DISCUSSION

3.1: SY-5609 ChIP-seq Experiments

Two biological ChIP-Seq replicates were generated from these experiments in HCT116 cells. Samples were treated with either DMSO, SY-5609 (Figure 9B), KB-0742 (Figure 9A) or both inhibitors. ChIP was performed using antibodies for pSer2, pSer5, and pSer7 on the RNAPII CTD, along with the RNAPII NTD as an estimate of total RNAPII. Additionally, one beads-only (no antibody) IP was performed to assess background binding (Figure S11), but was not sequenced. Sequencing this sample for subsequent analyses could provide information about sites of read accumulation in the absence of any enrichment, which could serve as a valuable reference point for enriched samples.

A B

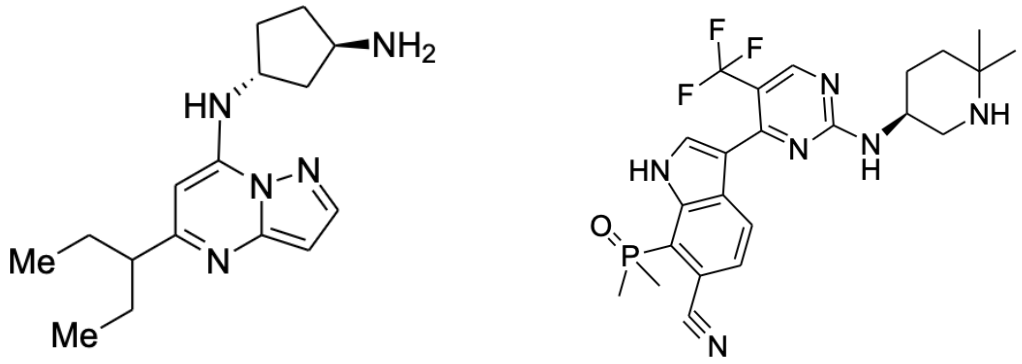


Figure 9: Structure of kinase inhibitors A) KB-0742 and B) SY-5609. KB-0742 inhibits CDK9 and SY-5609 inhibits CDK7. Structures adapted from Clopper and Taatjes 2019.

Replicate 1 displayed low read alignments, high duplication rates, and low mapped reads overall. Read quality declined drastically in both replicates after the 60-70 bp mark. (Figure 10A). To account for this, I used cutadapt to trim each sequence down to 60 bp , and reran trimmed samples through my quality control and read mapping pipeline to see if overall statistics improved. While trimming the reads resolved quality errors and yielded

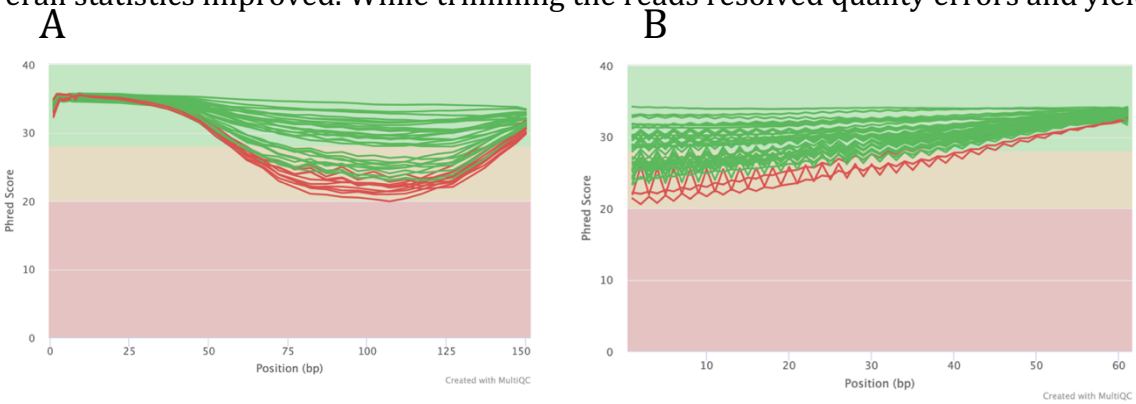


Figure 10: Phred Scores of Replicate Before (A) and After (B) Trimming With Cutadapt. Read quality of raw fastq files was assessed using FastQC version 0.11.8 and compiled into interactive plots using MultiQC v1.15. Cutdadapt version 4.4 was used to trim 90 bp off the 3' end of all reads to improve quality. Each line in the graph represents a sample. Red lines indicate samples whose Phred scores are too low to pass QC thresholds. Green lines indicate samples that pass QC thresholds.

higher overall Phred scores (Figure 10B) , it did not improve the percentage of reads mapped to the genome . Because of this, Replicate 1 was not analyzed further.

Replicate 2 was sequenced twice to attain the desired read depth. I combined fastq files from corresponding samples to generate pooled samples from both runs. Initial analysis of fastqc data showed highly irregular sequencing depths between IPs despite attempts to sequence with a 5:3:1 ratio of total RNAPII: RNAPII phosphomarks: Input DNA (Figure S27-S28). Of note, Serine 5 and total RNAPII datasets displayed relatively low read depths and Serine 2 and Serine 7 datasets displayed relatively high read depths. Input samples sequenced at a very low depth, and were not used for data normalization.

Metagene profiles were generated for the TSS and TES of genes from two datasets. The first dataset contained genes typically upregulated during the interferon response[1] (n = 138). These genes were sourced from two subsets: GSEA[160] and RNA-Seq experiments

published by the Taatjes, Dowell, and Kovarik labs in 2019[161] (Table S30). Genes from GSEA (n = 55) were first discovered using oligonucleotide microarrays for 6,800 genes[160] and later verified as statistically significant using Gene Set Enrichment Analysis. Genes from RNA-Seq (n =83, padj < 0.1, lfc > 1) were generated by treating HCT116 cells with IFNγ or PBS and used gene body counts (+301 to end) to quantify differential expression[161]. The second dataset contains genes highly expressed in HCT116 cells[163] (n = 263) relative to 1084 other cancer cell lines sourced from microarray mRNA expression profiles published in the Broad Institute's Cancer Cell Line Encyclopedia[162] (Table S29.) Significance was determined by computing a Standardized Value that indicates relative functional associations. The absolute value of the Standarized Value is equal to the negative log base ten of p values from continuous gene expression datasets. Utilizing these gene sets allows us to draw conclusions based on trends between genes that respond to stimuli and genes that are basally expressed in HCT116 cells. That said, future analyses of this data will utilize larger sets of expressed genes in HCT116 cells to better represent trends in global gene expression.

Genome browser snapshots were generated at the *IRF1, ELK4*, *TRIB1,* and *PDP1* loci. *IRF1* and *ELK4* are highly upregulated in response to interferon stimulation (Figure S31); *TRIB1* and *PDP1* are highly expressed in HCT116 cells (Figure S30) and are not responsive to IFN. As such, the former two loci serve as an example of interferon responsive genes (IRGs), while the latter represent genes that are well expressed in HCT116 cells independent of IFN stimulation.

In the total RNAPII (NTD) immunoprecipitation, both plots show a strikingly high amount of RNAPII occupancy at both the TSS and TES in cells treated with KB-0742, a CDK9 inhibitor, when compared to other conditions (Figure 11). The TSS peak suggests an accumulation of paused polymerases caused by CDK9 inhibition. Cells treated with CDK7 inhibitor SY5609 also show a higher RNAPII occupancy when compared to vehicle-treated cells at both the TSS and TES. An accumulation of RNAPII at the TSS aligns with previous

reports of CDK7 inhibition from the Taatjes Lab[69], suggesting that promoter escape or potentially pause release are impaired[71,94,97]. When cells were treated with both inhibitors, a lower overall RNAPII occupancy was observed (Figure 11-13). This could suggest i) premature termination at gene 5'-ends, ii) reduced RNAPII loading at the PIC, and/or iii) a lack of RNAPII reinitiation in the absence of CDK7 and CDK9 function. Peaks at the TES mirror those at the TSS, which may simply reflect the overall reduction of RNAPII complexes escaping the pause site and entering gene bodies during productive elongation. Both the IFN-induced and basal expressed datasets showed similar trends, which implies little difference between the mechanisms of regulation between genes activated by stimuli and those that are not.

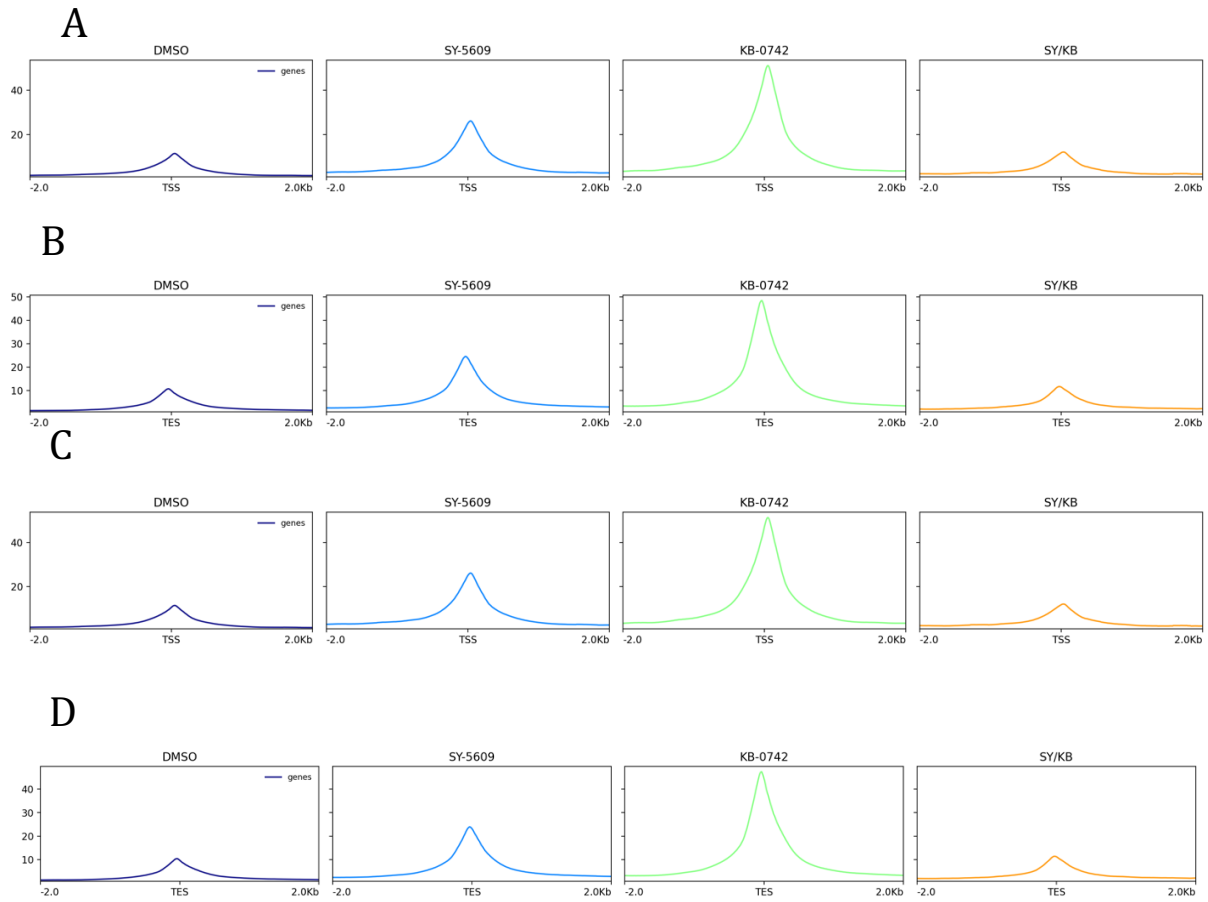


Figure 11: Metagene analysis profiles for RNAPII NTD ChIP-Seq. A,B) Traces mapped to an IRG dataset (n=138) at the TSS (A) and TES (B). C,D) Traces at the most highly expressed genes in HCT116 cells (n=263) at the TSS (C) and TES (D). All samples were normalized to 1x read coverage prior to metagene analysis. Read duplications, multimapped reads, and blacklisted genomic regions were filtered out during analysis. "SY/KB" is samples treated with both SY-5609 and KB-0742. Note that the SY/KB sample in this IP was subject to a reconditioning PCR cycle during the library preparation due to overamplification.

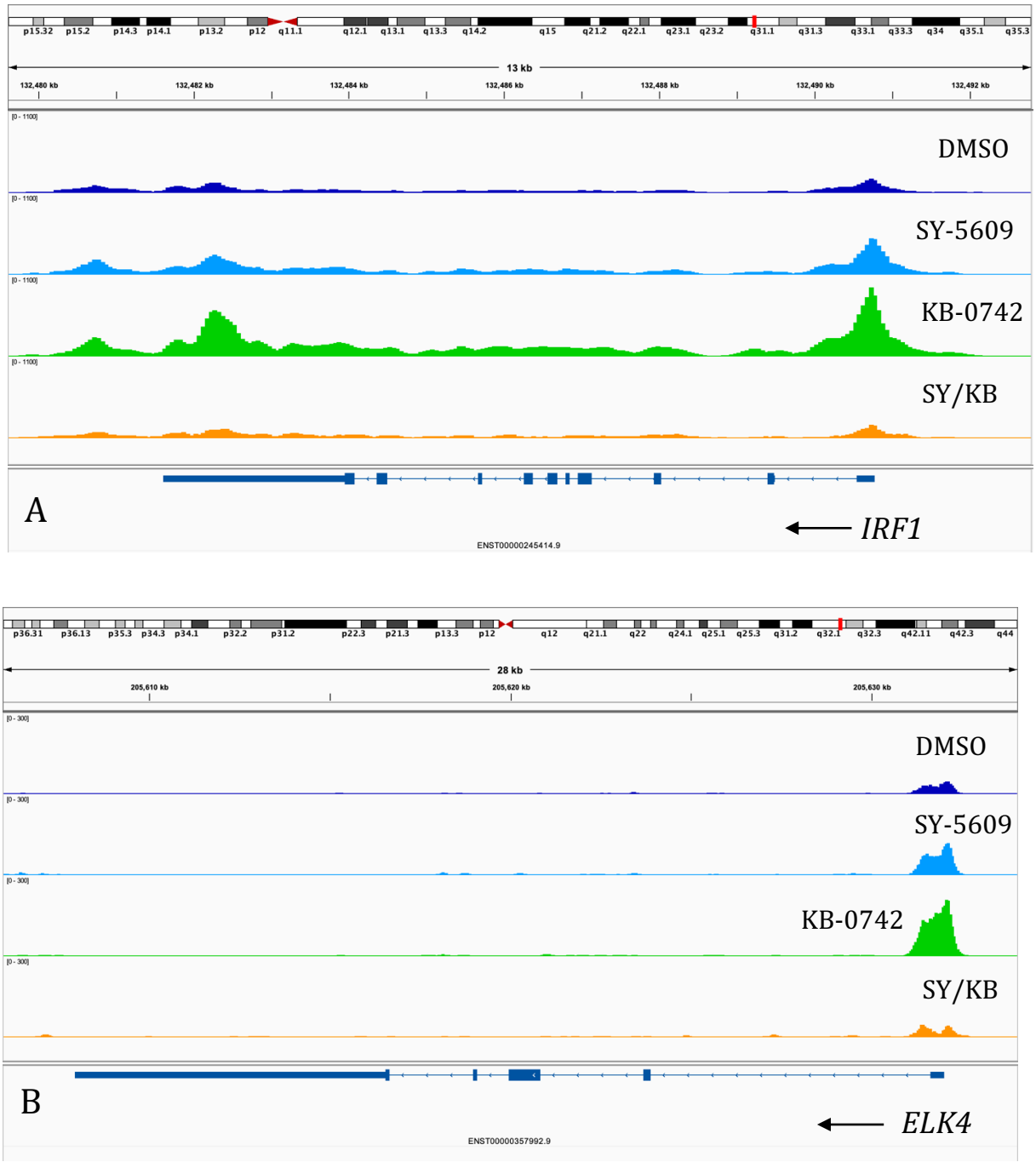
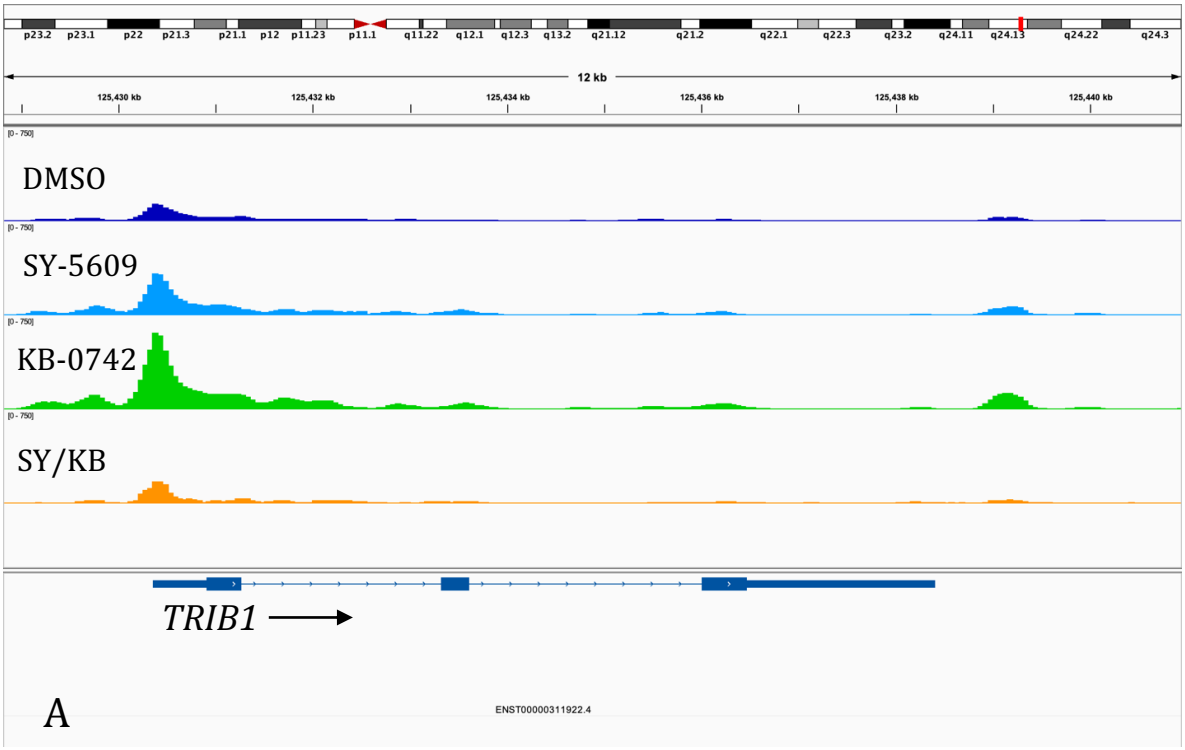
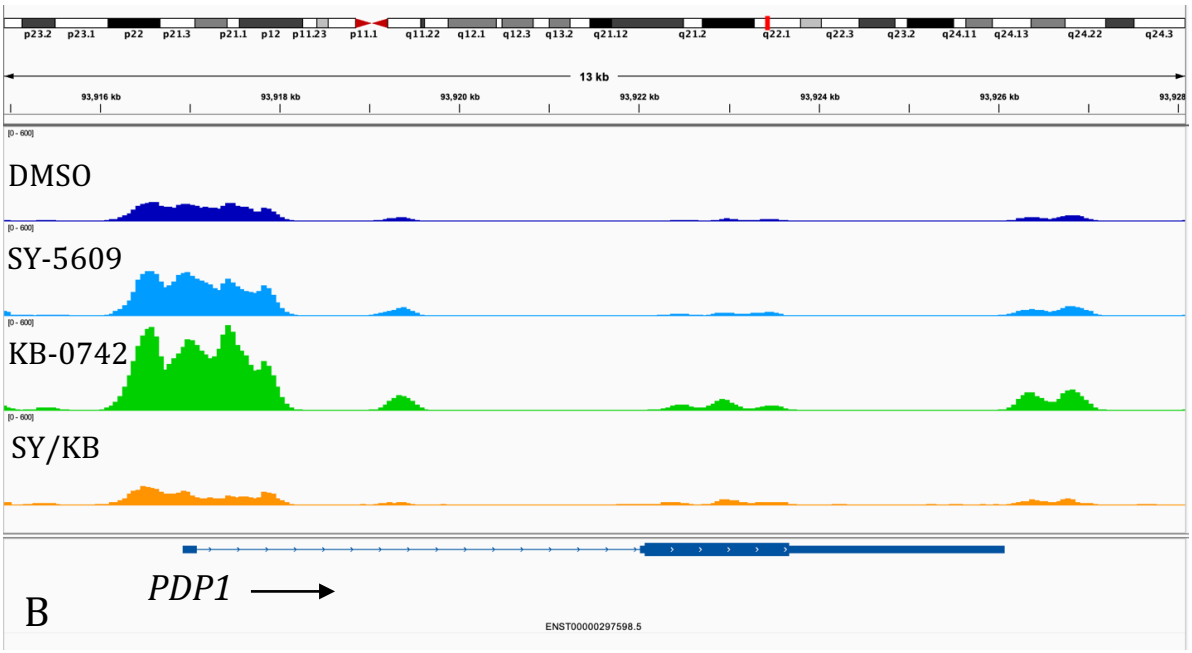


Figure 12: IGV snapshots of RNAPII NTD traces at *IRF1* (A) and *ELK4* (B). Genes are labeled at their 5' ends. Samples were normalized to 1x read coverage and mapped to the hg38 RefSeq MANE genome annotation to limit isoform variants. A) IGV scale = 0-1100. B) IGV scale = 0-300. Signals were not represented on a log scale.





<u>Figure 13: IGV snapshots of RNAPII NTD traces at *TRIB1* (A) and *PDP1* (B).</u>
Genes are labeled at their 5' ends. Samples were normalized to 1x read coverage and mapped to the hg38 RefSeq MANE genome annotation to limit isoform variants. A) IGV scale = 0-750. B) IGV scale = 0-600. Signals were not represented on a log scale.

RNAPII CTD Ser2 phosphorylation

pSer2 levels appeared lowly enriched but relatively consistent between the TSS and TES across all conditions (Figure 14-16). This stands in contrast to the conventional understanding of the distribution of this modification, as they are typically portrayed almost exclusively at or downstream of the TES[135,141,164]. That being said, pSer2 accumulation at the TSS has been reported by Schwartz et al., in which roughly 4,000 genes that display this phosphorylation pattern[138,139]. Further study is needed to determine if the genes identified by Schwartz et al. overlap with those identified in this study, and if there is any common biological pathway represented by these genes. Early research involving the human RNAPII CTD also identified this phenomenon[165], and a recent preprint from the Lis Lab observed pSer2 in the late pause region (30-60 nucleotides downstream of the TSS) using the novel method PRO-IP-Seq[139]. Despite this repeated observation, the role of pSer2 at the 5' end of RNAPII-transcribed genes is entirely uncharacterized.

Conditions in which CDK9 is inhibited (+KB0742) show an increase in pSer2 signal compared to vehicle (Figure 14). It is therefore possible that other kinases, such as CDK12/13, are compensating for CDK9. It is also possible that CDK9 inhibition results in dysregulation of phosphatases, causing any pSer2 marks that are deposited to remain on the CTD. CDK7 inhibition by SY5609 showed a similar impact, which may reflect its role as a CAK. Genes used in both the IFNγ and HCT116 datasets were not filtered by their proximity to other genes, which creates the potential for TSS and TES overlap. Future analysis of these datasets will account for this to avoid confounding signals.

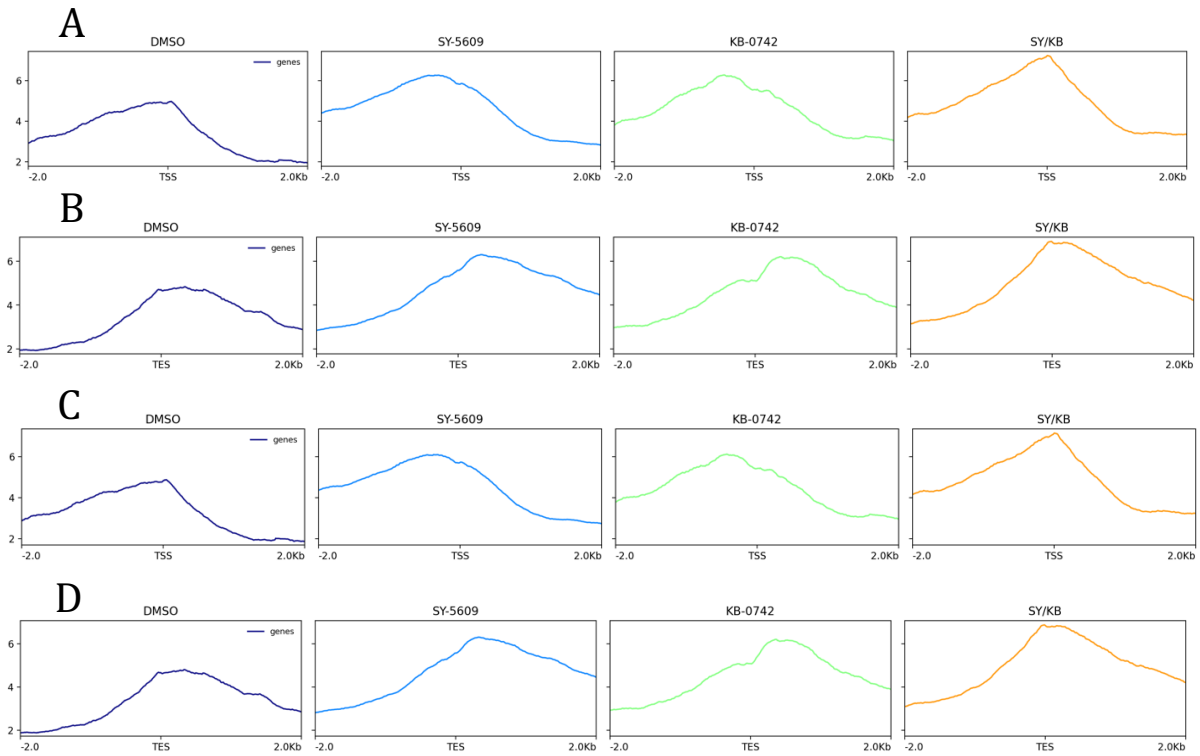


Figure 14: Metagene analysis profiles for RNAPII CTD pSer2 ChIP-Seq. A, B) Traces mapped to an IRG dataset (n=138) at the TSS (A) and TES (B). C, D) Traces at the most highly expressed genes in HCT116 cells (n=263) at the TSS (C) and TES (D). All samples were normalized to 1x coverage prior to metagene analysis. Read duplications and blacklisted genomic regions were removed. "SY/KB" is samples treated with both SY-5609 and KB-0742. Note that the KB-0742 sample in this IP was subject to a reconditioning PCR cycle during the library preparation due to overamplification.

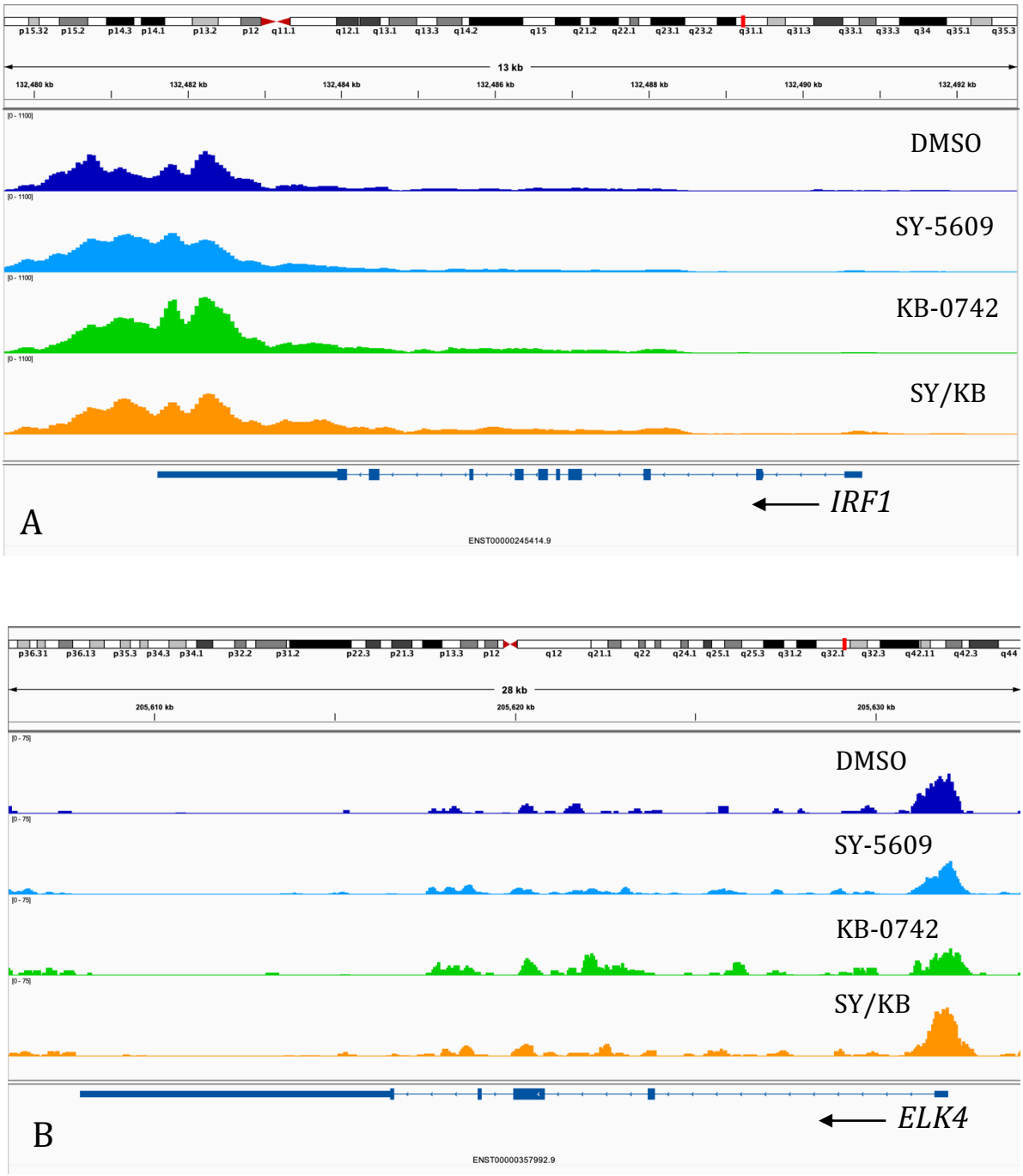
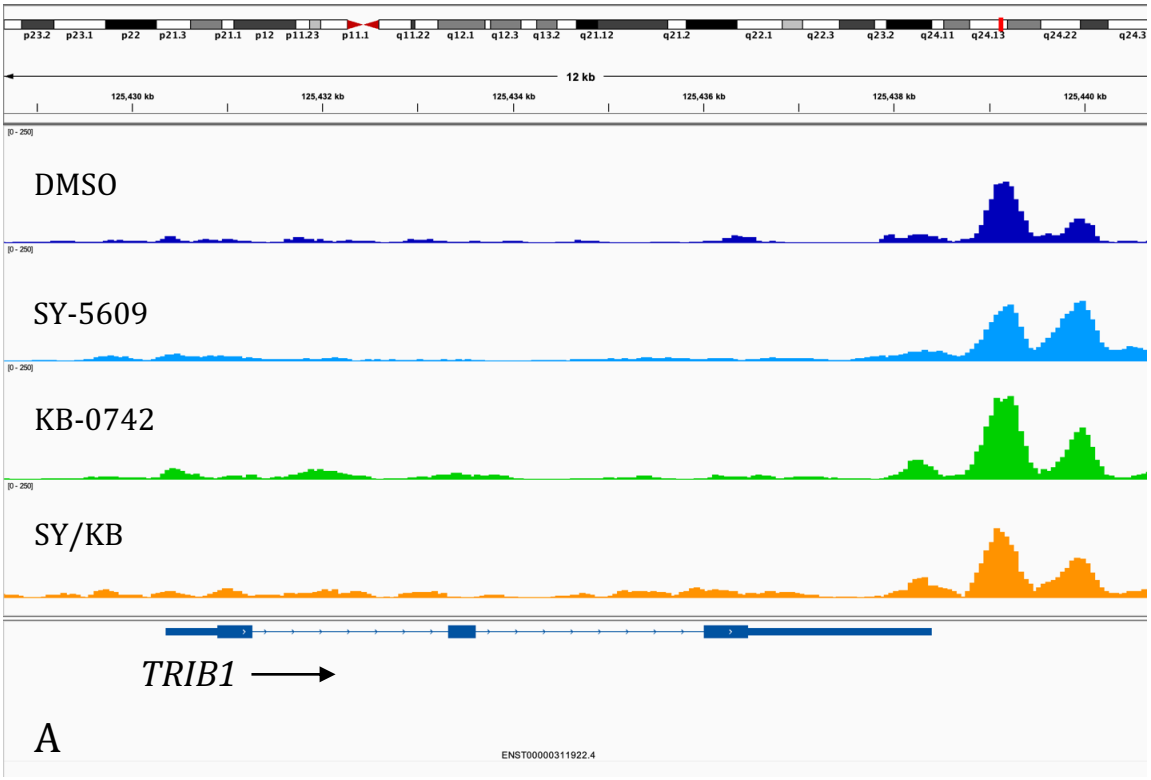


Figure 15: Representative genome viewer snapshots of RNAPII pSer2 traces at *IRF1* (A) and *ELK4* (B). Genes are labeled at their 5' ends. Samples were normalized to 1x read coverage and mapped to the hg38 RefSeq MANE genome annotation to limit isoform variants. A) IGV scale = 0-1100. B) IGV scale = 0-75. Signals are not represented on a log scale.



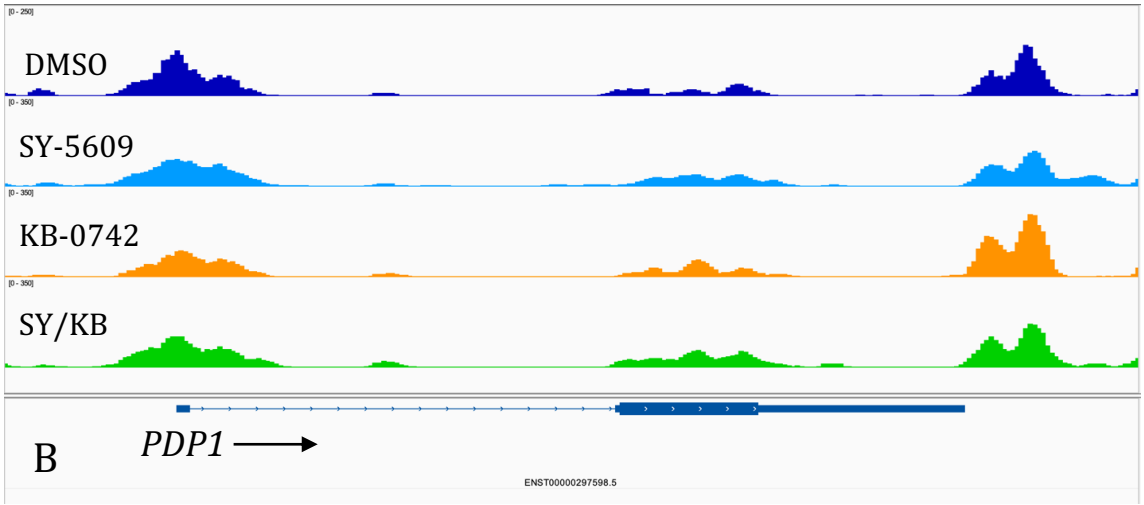


Figure 16: Representative genome viewer snapshots of RNAPII pSer2 traces at *TRIB1* (A) and *PDP1* (B). Genes are labeled at their 5' ends. Samples were normalized to 1x read coverage and mapped to the hg38 RefSeq MANE genome annotation to limit isoform variants. A) IGV scale = 0-250. B) IGV scale = 0-250. Signals are not represented on a log scale.

RNAPII CTD Ser5 phosphorylation

Immunoprecipitations specific for the Serine 5 phosphorylated RNAPII showed an increase in read density around the TSS in all conditions relative to the vehicle treated cells (Figure 17). This again potentially reflects CDK7 and CDK9 roles in pausing. Both sets of genes (IFN-induced and basal expressed) showed similar enrichment, suggesting that genes activated by cell signaling pathways during a stimulus are not differentially affected by CDK7 or CDK9, at least with respect to RNAPII CTD phosphorylation. That being said, the SY-5609 and dual inhibitor groups displayed "towering" in the genome browser snapshots, reflecting their low read coverage and a high amount of duplication, which should be taken into account during the interpretation of these datasets (Figure 18-19).

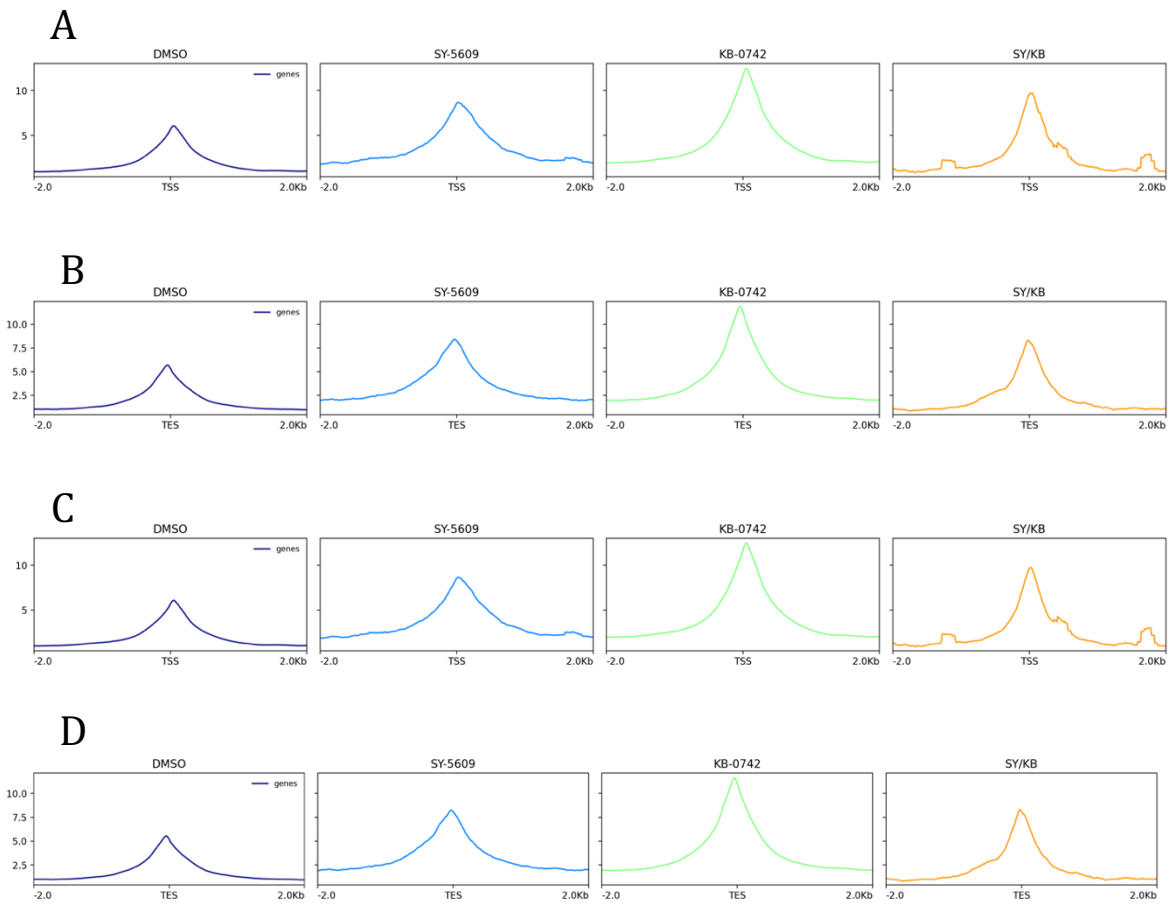


Figure 17: Metagene analysis profiles for RNAPII CTD pSer5 ChIP-Seq. A,B) Traces mapped to an IRG dataset (n=138) B, C) Traces at the most highly expressed genes in HCT116 cells (n=263). The TSS was used as a reference point and reads were not scaled to equivalent lengths. All samples were normalized to 1x read coverage prior to metagene analysis. Read duplications and blacklisted genomic regions were filtered out during analysis. "SY/KB" is samples treated with both SY-5609 and KB-0742. Note that the DMSO and SY5609 samples in this IP were subject to a reconditioning PCR cycle during the library preparation due to overamplification.

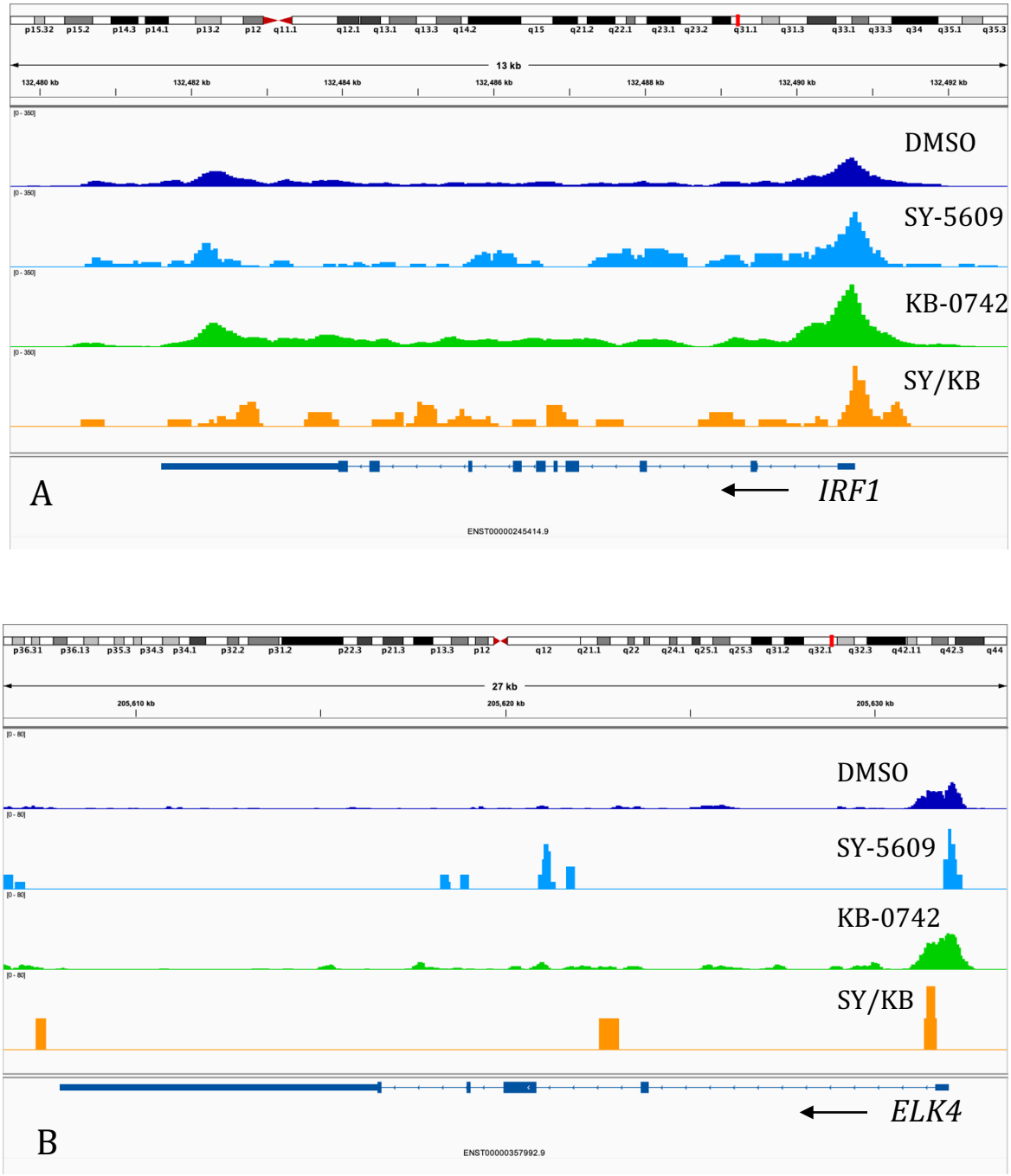


Figure 18: Representative genome viewer snapshots of RNAPII pSer5 traces at *IRF1* (A) and *ELK4* (B). Genes are labeled at their 5' ends. Samples were normalized to 1x read coverage and mapped to the hg38 RefSeq MANE genome annotation to limit isoform variants. A) IGV scale = 0-350. B) IGV scale = 0-80. Signals are not represented on a log scale.

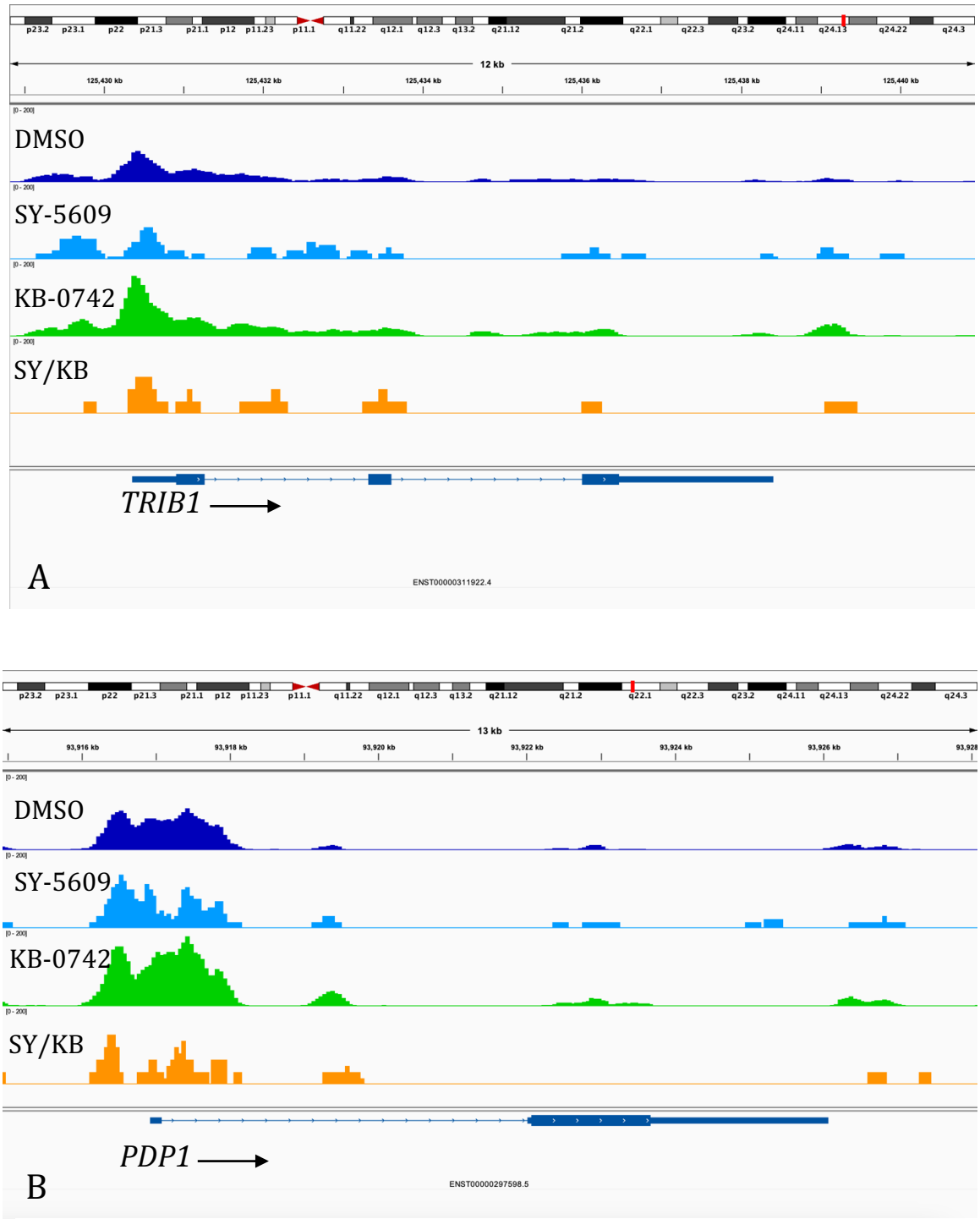
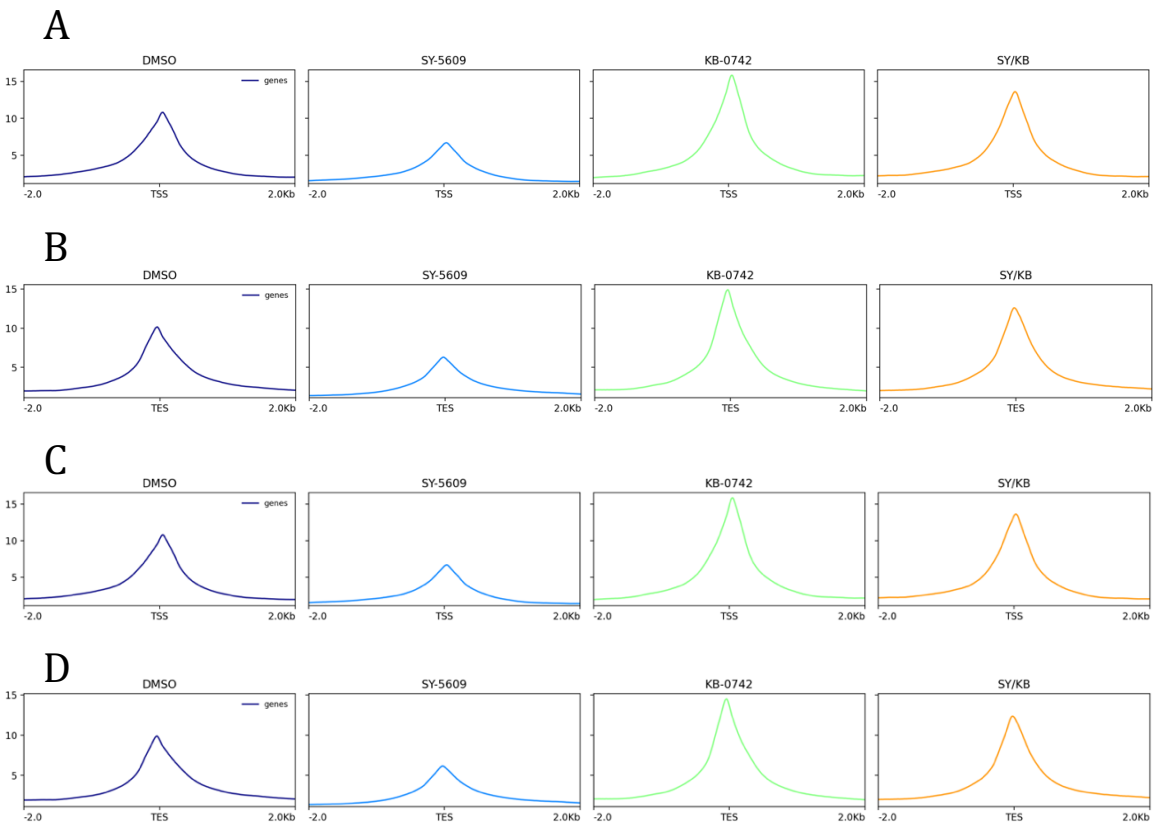


Figure 19: Representative genome viewer snapshots of RNAPII pSer5 traces at *TRIB1* (A) and *PDP1* (B). Genes are labeled at their 5' ends. Samples were normalized to 1x read coverage and mapped to the hg38 RefSeq MANE genome annotation to limit isoform variants. A) IGV scale = 0-200. B) IGV scale = 0-200. Signals are not represented on a log scale.

RNAPII CTD Ser7 phosphorylation

pSer7 marks are deposited by CDK7[69,166], and are typically expected to peak near the TSS and gradually increase throughout the gene body. In both the interferon and basal HCT116 datasets, peaks at the TSS and TES emerged in all four experimental conditions (Figure 20-22). However, SY-5609 treatment appeared to decrease enrichment of pSer7 at both the TSS and TES when compared to both the DMSO-treated control group and the KB-0742 treated group. This is in line with our understanding of CDK7 as the primary depositor of serine 7 phosphomarks[69,166]. The CDK9 inhibited profile shows greater enrichment than the vehicle at the TSS, which again coincides with our understanding of CDK9 as the major catalyst for pause release. The dual inhibitor condition appears to create an intermediate enrichment pattern that falls between the SY-5609 and KB-0742 groups, suggesting that CDK9 acts as an antagonist to CDK7 with respect to pSer7 phosphorylation.

Overall, SY5609 appears to decrease pSer7 marks as a function of CDK7 inhibition. This could be due to the fact that pSer7 serves an auxiliary function for which compensation may be less likely than pSer5, which does not respond in this manner to SY-5609 treatment (Figure 15, 16), despite also being a CDK7 target. Subsequent experiments are needed to further test this observation.



<u>Figure 20: Metagene analysis profiles for RNAPII CTD pSer7 ChIP-Seq</u>. A.) Traces mapped to an IRG dataset (n=138) B.) Traces at the most highly expressed genes in HCT116 cells (n=263). The TSS was used as a reference point and reads were not scaled to equivalent lengths. All samples were normalized to 1x read coverage prior to metagene analysis. Read duplications and blacklisted genomic regions were removed during analysis.

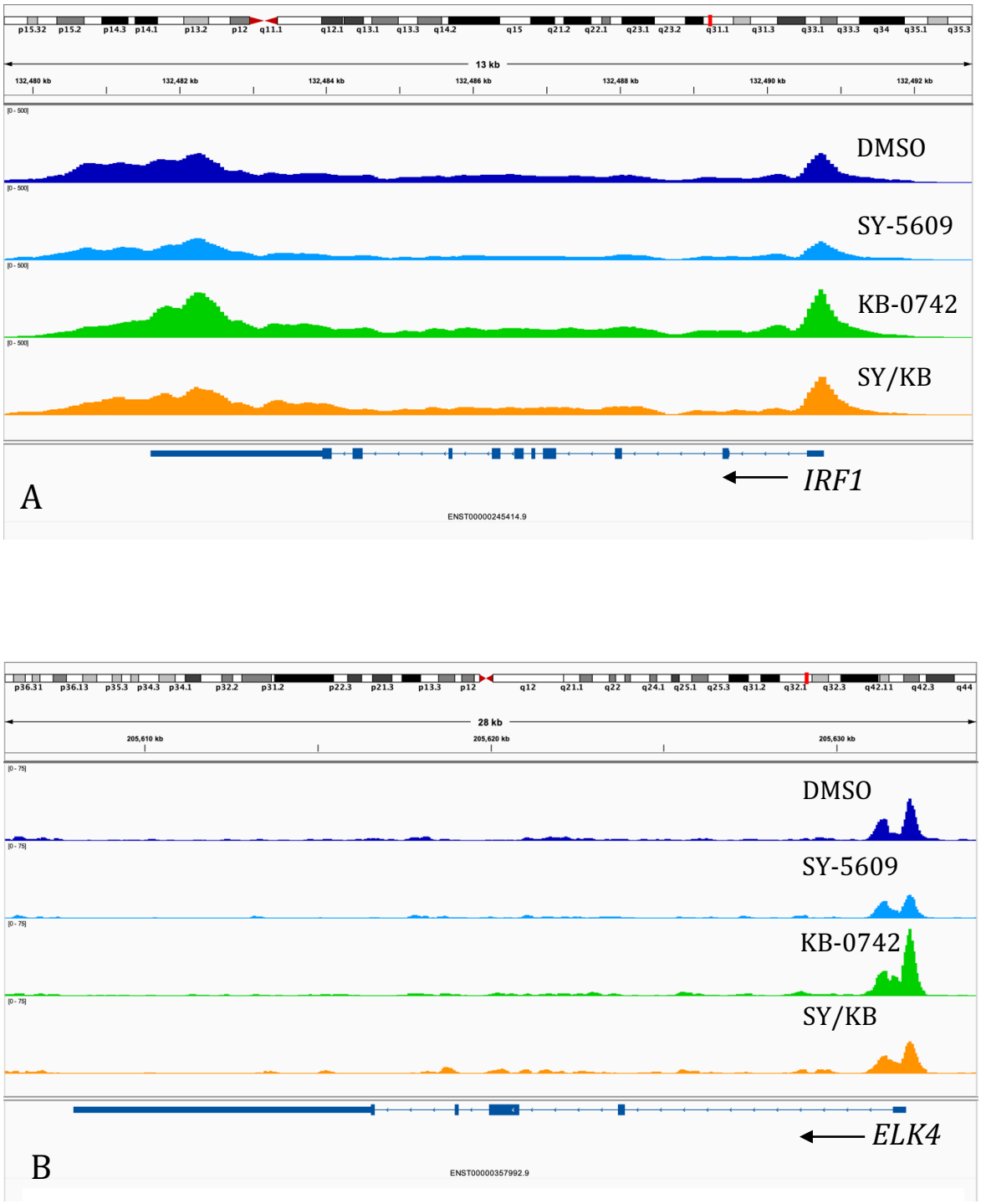
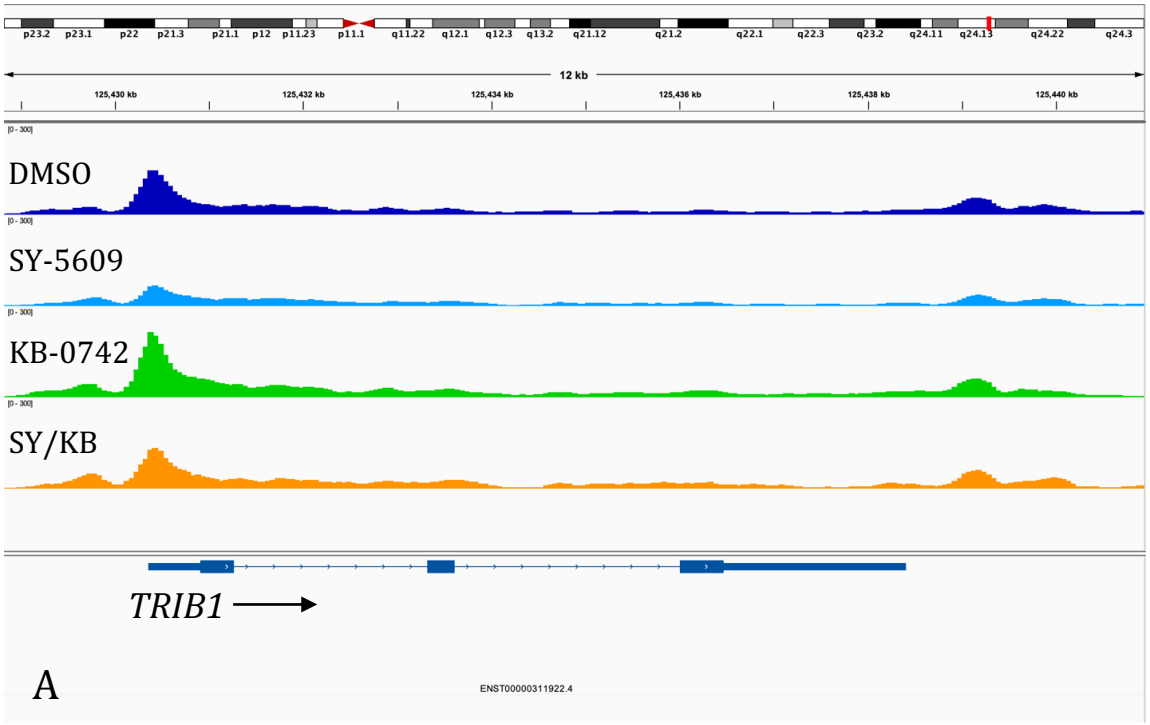


Figure 21: Representative genome viewer snapshots of RNAPII pSer7 traces at *IRF1* (A) and *ELK4* (B). Genes are labeled at their 5' ends. Samples were normalized to 1x read coverage and mapped to the hg38 RefSeq MANE genome annotation to limit isoform variants. A) IGV scale = 0-500. B) IGV scale = 0-75. Signals are not represented on a log scale.



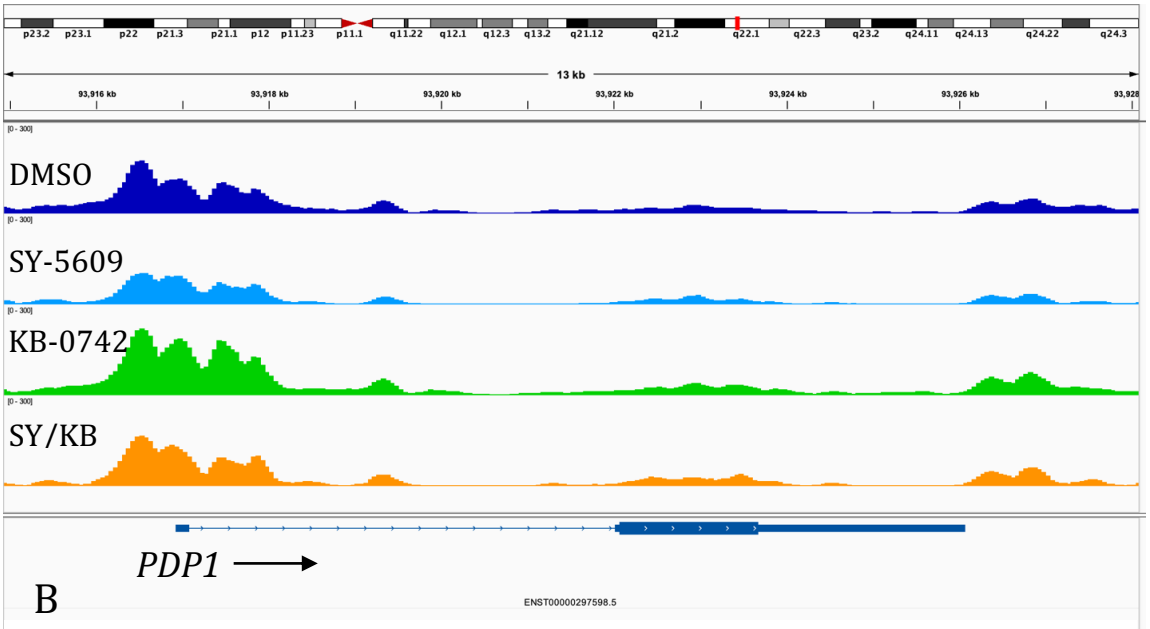


Figure 22: Representative genome viewer snapshots of RNAPII pSer7 traces at *TRIB1* (A) and *PDP1* (B). Genes are labeled at their 5' ends. Samples were normalized to 1x read coverage and mapped to the hg38 RefSeq MANE genome annotation to limit isoform variants. A) IGV scale = 0-300. B) IGV scale = 0-300. Signals are not represented on a log scale.

Overall, this data displayed enrichment patterns that both answered and raised new questions about CDK7 and CDK9's regulation of transcription. In every IP, cells treated with only CDK9 inhibitor (KB0742) showed consistently high peaks at the TSS when compared to their counterparts. This likely reflects the impairment of pause release. DMSO-treated cells often displayed the lowest enrichment across all samples. In TSS profiles, this suggests that transcription proceeds through the pause checkpoint into productive elongation without error, as is the case when both CDK7 and CDK9 kinases are active. The peaks observed in TES profiles are likely in part a function of slower RNAPII speeds in this region. There were also few differences between IRGs and genes highly expressed in HCT116 cells, suggesting that the impact of CDK7 and CDK9 inhibition is not overridden by stimulus response pathways, at least not in the context of the RNAPII CTD.

The most compelling aspect of this data lies in the phosphorylation trends observed in Serine 2 and Serine 7. In yeast, pSer2 is not observed at gene 5'-ends. Instead, its levels increase gradually across the gene body, and peak at the 3'-end[141,164,167,168]. Serine 2 phosphorylation was readily observed at the 5' end of genes, which is a pattern that appears to have been documented in the literature but is rarely discussed or included in standard models. Clearly, more extensive study is needed to determine what role pSer2 might play at the 5' ends of genes.

Serine 7 phosphorylation was decreased in the presence of SY5609 but not in the SY/KB group, suggesting that SY5609 has a negative effect on pSer7 marks that is not present when CDK9 is also inhibited. This is in contrast to what has been observed in similar experiments in yeast, which show an equal decrease in pSer7 at the TSS in both *kin28-as* and *kin28-as/Δbur2* groups[168], which are equivalent to SY-5609 and SY/KB treatments respectively. These studies also showed that *Δbur2* cells have roughly equal amount of pSer7 compared to WT at the TSS with minimal change in total RNAPII between groups[168], which may suggest that pSer7 levels are not amplified by an accumulation of RNAPII at the TSS during CDK9 inhibition in yeast . This may implicate CDK9 as an

antagonist to CDK7, which may function as a form of coregulation between these kinases. It is also possible that transcriptional CDKs display a level of compensation, which might involve additional kinases such as CDK12 and CDK13.

Some of these samples, namely in the Serine 5 data, also exhibited high duplication rates after sequencing. While duplicates were removed during the metagene analysis, this likely impacted the complexity of these dataset and limited the results to read depths that likely do not provide the full scope of pSer5 phosphorylation trends in response to CDK7 and CDK9 inhibition.

It is also worth noting that genes in these datasets were not filtered by their proximity to other genes. This could generate confounding signals between the TSS of one gene and the TES of another, which may be in part why we observed unexpected peaks at the TES. Future analysis of this data will utilize exclusively genes that are at least five kilobases away from other genes to help reduce noise at target genes generated by proximal signals.

3.2: SY-351

Prior to performing the SY-5609 experiments, a similar experimental scheme was employed using SY-351, a covalent inhibitor for CDK7 (Figure 23, Figure S39). SY-351 proved unsuitable for PRO-Seq experiments performed by other Taatjes lab members as it caused a high level of global transcriptional downregulation that impeded normalization during differential expression analysis. Because of this, we switched the main ChIP-Seq experimental setup to utilize the previously discussed non-covalent CDK7 inhibitor SY-5609 to complement the PRO-Seq datasets. However, we also sequenced two ChIP-Seq technical replicates from original SY-351 treated groups. Libraries were demultiplexed, and quality was assessed via

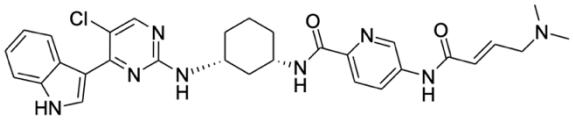


Figure 23: Structure of SY-351. Structure adapted from Clopper and Taatjes 2019.

Q-scores. Samples enriched for the RNAPII NTD were sequenced to a target depth of 50 million reads per library, while pSer5 phosphomarks were sequenced to a target depth of 30 million reads. All input samples were sequenced to a target depth of 10 million reads. For the NTD and pSer5 enriched samples, our target variance was +/- 10 million reads. Overall, the quality of reads returned from sequencing was high (Figures S23-S25).

Metaplots that mapped read coverage at interferon responsive genes (IRGs)[160] showed modest enrichment and little change between conditions (Figure 24). pSer5 RNAPII appears to represent a subset of total RNAPII as predicted. pSer5 marks appeared to decrease in the presence of SY-351, a CDK7 inhibitor. However, the second dual inhibitor condition displayed a markedly different pattern, with a larger proportion of RNAPII possessing pSer5. Furthermore, the second technical replicate suggests that nearly all of RNAPII present in the inhibitor conditions possessed pSer5 marks, which was an unanticipated result.

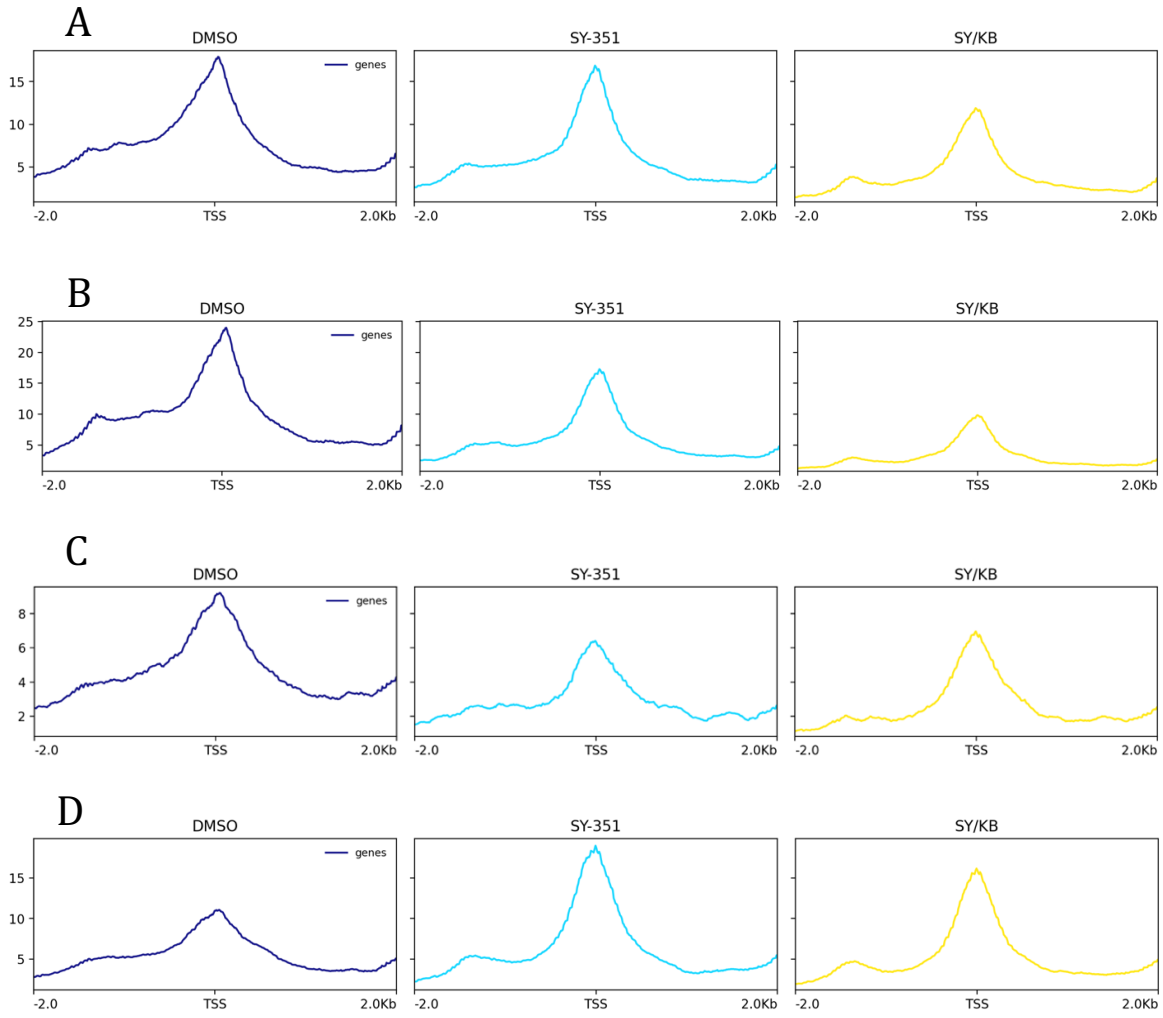


Figure 24: Metagene analysis profiles at IRG's for RNAPII NTD and pSer5 ChIP-Seq. All reads were mapped to an IRG dataset (n=71). A, B correspond to RNAPII NTD. C,D correspond to RNAPII pSer5. A,C were produced from Replicate 1. B,D were produced from Replicate 2. All samples were normalized to 1x read coverage prior to metagene analysis. Read duplications and blacklisted genomic regions were removed during analysis.

With this in mind, I sought to see how CDK7 and CDK9 inhibition impacted RNAPII phosphorylation and occupancy outside of a stimulus context. Reads were mapped to to a subset of genes calculated to be highly expressed in HCT116 cells[163] (Figure 25, Figure S29). This dataset produced similar trends to those observed in the interferon-responsive gene dataset, suggesting minimal difference between genes that are impacted by interferon stimulation and genes that are not.

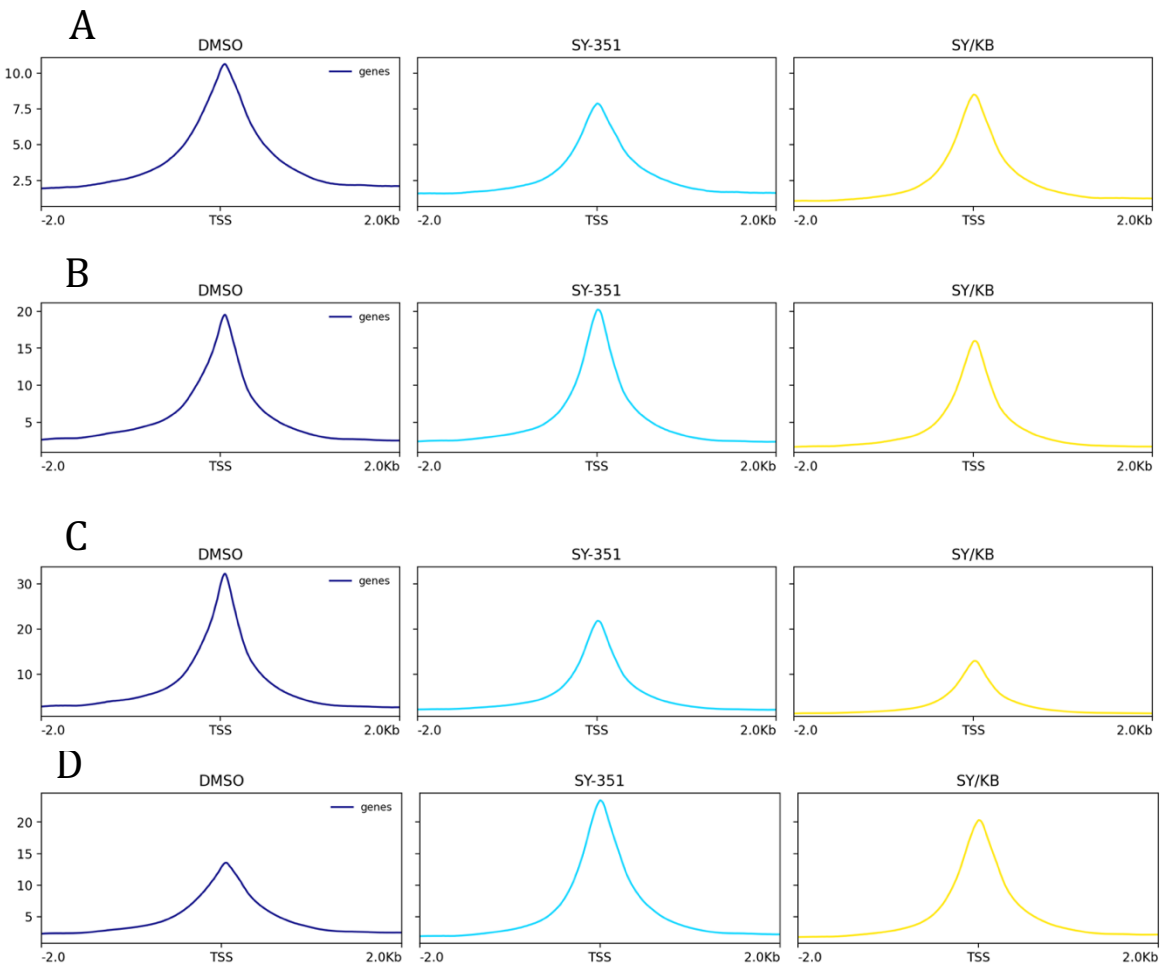


Figure 25: Metagene analysis profiles at highly expressed HCT116 genes for RNAPII NTD and pSer5 ChIP-Seq. All reads were mapped to the HCT116 dataset (n=263). A, B correspond to RNAPII NTD. C,D correspond to RNAPII pSer5. A,C were produced from Replicate 1. B,D were produced from Replicate 2. All samples were normalized to 1x read coverage prior to metagene analysis. Read duplications and blacklisted genomic regions were removed during analysis.

To explore this further, I sought to assess the efficacy of ChIP enrichment via an analysis of concordance between replicates. I performed two technical replicates of the SY-351 ChIP experiments. However, there was an error on the sequencer during the run of Technical Replicate 1, and as such these samples were sequenced twice to achieve the desired depth. To account for this, I combined corresponding samples from each run to produce a unified technical replicate. To gain an understanding of replicate concordance between replicates 1 and 2, I used the DeepTools plotCorrelation package to produce an assessment of similarity between samples via a spearman r coefficient. I chose to use this statistical test because it appropriately accounts for non-normal data distributions, which are common in ChIP-Seq. The ideal outcome from this analysis would be that identical samples from each replicate would display higher r coefficients than those with different experimental conditions in the same replicate, indicating that our replicates are statistically sound. However, I observed a relatively high spearman coefficients between nearly all samples, with the lowest being 0.70 and the highest being 0.96, barring those produced when samples were analyzed against themselves (Figure 26).

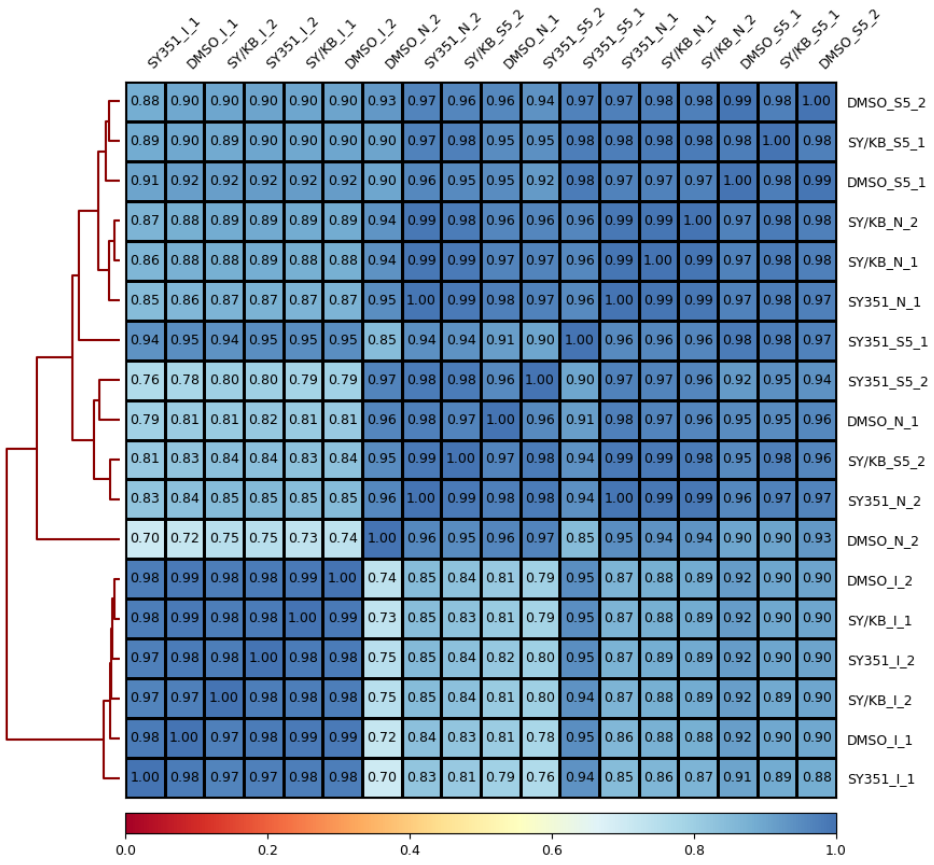


Figure 26: Spearmann's r coefficients calculated between replicates of SY-351 ChIP-Seq in HCT116 Genes (n = 263). Numbers 1 or 2 correspond to technical replicate ID's. "N" refers to RNAPII NTD ChIP samples, while "S5" refers to RNAPII pSer5 samples. "I" refers to DNA input. "SY/KB" denotes samples treated with SY-351 and KB-0742.

This analysis suggests little enrichment in these samples above background signal. I had anticipated that ChIP enrichment signals would vary both between inhibitor and control conditions and between interferon-response genes and the genome at large. I was unfortunately unable to make meaningful distinctions between these conditions due to low ChIP enrichment.

4. CONCLUDING REMARKS & FUTURE DIRECTIONS

These experiments sought to understand how CDK7 and CDK9 regulate transcription through their phosphorylation of the RNAPII CTD. They also attempted to gain a clearer picture of how CDK7 regulates 3' end processing and splicing events. Both of these events are poorly understood due to the limitations of early chemical inhibitors (e.g. flavopiridol, THZ1) and short-read sequencing technology. I had access to state of the art next-generation inhibitors for both CDK7 and CDK9, and used these to probe the regulatory impact of the kinases on RNAPII via ChIP-seq. I also performed the Taatjes lab's first long-read RNA sequencing experiment in collaboration with the Ramani lab (UCSF) to explore how inhibiting CDK7 impacts splicing events and 3' processing under normal conditions and in response to heat shock. I used these methods to answer fundamental questions about the highly orchestrated process of eukaryotic gene expression, which impacts an array of processes that are key to our understanding of both basic cellular function and a slew of diseases.

While these preliminary findings are intriguing and warrant further investigation, certain considerations must be taken into account when interpreting the data. Raw DNA yields from the ChIP-Seq DNA isolation were low (Figure S10, S11), pointing to the need for further experimental optimization during the ChIP process. This could be achieved by using less stringent wash buffers or changing the IP conditions. Additionally, due to the high volume of samples utilized in these experiments, most samples were stored at -80 °C following shearing. It is possible that the freezing of these samples in the shearing buffer, which contains a high percentage of SDS, negatively impacted the formaldehyde crosslinks and resulted in a loss of yield. The large volume of samples also generated inefficiency during IP setup and DNA isolation which may have also contributed to lower yields. Library preparation could be further optimized to avoid over-amplification of samples, which can reduce library complexity and result in high levels of duplications that decrease read depth. Additionally, these ChIP-seq experiments contained only one biological replicate. Future

experiments should include at least two biological replicates to ensure that any novel findings are not anomalous and can be tested using the appropriate statistics.

Cyclin-dependent kinases (CDKs) have intrigued researchers for decades due to their role in transcription regulation and their potential as therapeutic targets. This work used state-of-the-art inhibitors and NGS technology to probe the intricate regulatory mechanisms of CDK7 and CDK9 in human transcription, focusing on the RNAPII CTD. My work reaffirms our understanding of CDK9 as a major regulator of promoter-proximal RNAPII pausing and pause release, and additionally posits that CDK9 may act antagonistically toward CDK7. Serine 7 is implicated as being majorly affected by CDK7 inhibition, whereas Serine 5 surprisingly was not. This discovery suggests that CDK7-dependent effects on RNAPII activity, for example co-transcriptional splicing[117], might result from altered pSer7 rather than altered pSer5. pSer2 marks at gene 5' ends were also discovered to be widespread in these experiments. The role of pSer2 at gene 5'-ends is understudied, but appears to be specific to metazoans, as this pattern is not observed in yeast[141,164,167,168]. Phosphorylation trends also point to a level of kinase compensation or phosphatase dysregulation that warrants further exploration.

5. REFERENCES

1 Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F, *et al.* An estimation of the number of cells in the human body. *Annals of Human Biology* 2013;**40**:463–71. https://doi.org/10.3109/03014460.2013.807878.

2 Valentine JW, Collins AG, Meyer CP. Morphological complexity increase in metazoans. *Paleobiology* 1994;**20**:131–42. https://doi.org/10.1017/S0094837300012641.

3 Vickaryous MK, Hall BK. Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biol Rev Camb Philos Soc* 2006;**81**:425–55. https://doi.org/10.1017/S1464793106007068.

4 Roy AL, Conroy RS. Toward mapping the human body at a cellular resolution. *Mol Biol Cell* 2018;**29**:1779–85. https://doi.org/10.1091/mbc.E18-04-0260.

5 Crick F. Central Dogma of Molecular Biology. *Nature* 1970;**227**:561–3. https://doi.org/10.1038/227561a0.

6 Zhang P, Wu W, Chen Q, Chen M. Non-Coding RNAs and their Integrated Networks. *J Integr Bioinform* 2019;**16**:20190027. https://doi.org/10.1515/jib-2019-0027.

7 Girbig M, Misiaszek AD, Müller CW. Structural insights into nuclear transcription by eukaryotic DNA-dependent RNA polymerases. *Nat Rev Mol Cell Biol* 2022;**23**:603–22. https://doi.org/10.1038/s41580-022-00476-9.

8 Vo ngoc L, Wang Y-L, Kassavetis GA, Kadonaga JT. The punctilious RNA polymerase II core promoter. *Genes Dev* 2017;**31**:1289–301. https://doi.org/10.1101/gad.303149.117.

9 Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol* 2018;**19**:621–37. https://doi.org/10.1038/s41580-018-0028-8.

10 Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 2006;**38**:626–35. https://doi.org/10.1038/ng1789.

11 Panigrahi A, O'Malley BW. Mechanisms of enhancer action: the known and the unknown. *Genome Biology* 2021;**22**:108. https://doi.org/10.1186/s13059-021-02322-1.

12 Wang M, Chen Q, Wang S, Xie H, Liu J, Huang R, *et al.* Super-enhancers complexes zoom in transcription in cancer. *Journal of Experimental & Clinical Cancer Research* 2023;**42**:183. https://doi.org/10.1186/s13046-023-02763-5.

13 Popay TM, Dixon JR. Coming full circle: On the origin and evolution of the looping model for enhancer–promoter communication. *Journal of Biological Chemistry* 2022;**298**:. https://doi.org/10.1016/j.jbc.2022.102117.

14 Verrijzer CP, Chen J-L, Yokomori K, Tjian R. Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* 1995;**81**:1115–25. https://doi.org/10.1016/S0092-8674(05)80016-9.

15 Chen Z, Manley JL. Core Promoter Elements and TAFs Contribute to the Diversity of Transcriptional Activation in Vertebrates. *Mol Cell Biol* 2003;**23**:7350–62. https://doi.org/10.1128/MCB.23.20.7350-7362.2003.

16 Maston GA, Zhu LJ, Chamberlain L, Lin L, Fang M, Green MR. Non-canonical TAF complexes regulate active promoters in human embryonic stem cells. *eLife* 2012;**1**:e00068. https://doi.org/10.7554/eLife.00068.

17 Reinberg D, Orphanides G, Ebright R, Akoulitchev S, Carcamo J, Cho H, *et al.* The RNA polymerase II general transcription factors: past, present, and future. *Cold Spring Harb Symp Quant Biol* 1998;**63**:83–103. https://doi.org/10.1101/sqb.1998.63.83.
18 Schier AC, Taatjes DJ. Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev* 2020;**34**:465–88. https://doi.org/10.1101/gad.335679.119.
19 Farnung L, Vos SM. Assembly of RNA polymerase II transcription initiation complexes. *Curr Opin Struct Biol* 2022;**73**:102335. https://doi.org/10.1016/j.sbi.2022.102335.
20 Conaway RC, Conaway JW. General initiation factors for RNA polymerase II. *Annu Rev Biochem* 1993;**62**:161–90. https://doi.org/10.1146/annurev.bi.62.070193.001113.
21 Littlefield O, Korkhin Y, Sigler PB. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proceedings of the National Academy of Sciences* 1999;**96**:13668–73. https://doi.org/10.1073/pnas.96.24.13668.
22 Patel AB, Louder RK, Greber BJ, Grünberg S, Luo J, Fang J, *et al.* Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* 2018;**362**:eaau8872. https://doi.org/10.1126/science.aau8872.
23 Louder RK, He Y, López-Blanco JR, Fang J, Chacón P, Nogales E. Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* 2016;**531**:604–9. https://doi.org/10.1038/nature17394.
24 Cianfrocco MA, Kassavetis GA, Grob P, Fang J, Juven-Gershon T, Kadonaga JT, *et al.* Human TFIID binds to core promoter DNA in a reorganized structural state. *Cell* 2013;**152**:120–31. https://doi.org/10.1016/j.cell.2012.12.005.
25 Liu X, Bushnell DA, Wang D, Calero G, Kornberg RD. Structure of an RNA Polymerase II–TFIIB Complex and the Transcription Initiation Mechanism. *Science* 2010;**327**:206–9. https://doi.org/10.1126/science.1182015.
26 Imbalzano AN, Zaret KS, Kingston RE. Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. *Journal of Biological Chemistry* 1994;**269**:8280–6. https://doi.org/10.1016/S0021-9258(17)37190-9.
27 Kostrewa D, Zeller ME, Armache K-J, Seizl M, Leike K, Thomm M, *et al.* RNA polymerase II–TFIIB structure and mechanism of transcription initiation. *Nature* 2009;**462**:323–30. https://doi.org/10.1038/nature08548.
28 Ohkuma Y, Roeder RG. Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. *Nature* 1994;**368**:160–3. https://doi.org/10.1038/368160a0.
29 Watanabe T, Hayashi K, Tanaka A, Furumoto T, Hanaoka F, Ohkuma Y. The Carboxy Terminus of the Small Subunit of TFIIE Regulates the Transition from Transcription Initiation to Elongation by RNA Polymerase II. *Mol Cell Biol* 2003;**23**:2914–26. https://doi.org/10.1128/MCB.23.8.2914-2926.2003.
30 Lin YC, Gralla JD. Stimulation of the XPB ATP-dependent helicase by the beta subunit of TFIIE. *Nucleic Acids Research* 2005;**33**:3072–81. https://doi.org/10.1093/nar/gki623.
31 Kim Y-J, Björklund S, Li Y, Sayre MH, Kornberg RD. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 1994;**77**:599–608. https://doi.org/10.1016/0092-8674(94)90221-6.

32 Jiang YW, Veschambre P, Erdjument-Bromage H, Tempst P, Conaway JW, Conaway RC, *et al.* Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. *Proceedings of the National Academy of Sciences* 1998;**95**:8538–43. https://doi.org/10.1073/pnas.95.15.8538.
33 Meyer KD, Lin S, Bernecky C, Gao Y, Taatjes DJ. p53 activates transcription by directing structural shifts in Mediator. *Nat Struct Mol Biol* 2010;**17**:753–60. https://doi.org/10.1038/nsmb.1816.
34 Richter WF, Nayak S, Iwasa J, Taatjes DJ. The Mediator complex as a master regulator of transcription by RNA polymerase II. *Nat Rev Mol Cell Biol* 2022:1–18. https://doi.org/10.1038/s41580-022-00498-3.
35 Allen BL, Taatjes DJ. The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol* 2015;**16**:155–66. https://doi.org/10.1038/nrm3951.
36 Jeronimo C, Langelier M-F, Bataille AR, Pascal JM, Pugh BF, Robert F. Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. *Molecular Cell* 2016;**64**:455–66. https://doi.org/10.1016/j.molcel.2016.09.002.
37 Petrenko N, Jin Y, Wong KH, Struhl K. Mediator Undergoes a Compositional Change during Transcriptional Activation. *Molecular Cell* 2016;**64**:443–54. https://doi.org/10.1016/j.molcel.2016.09.015.
38 Ramasamy S, Aljahani A, Karpinska MA, Cao TBN, Velychko T, Cruz JN, *et al.* The Mediator complex regulates enhancer-promoter interactions. *Nat Struct Mol Biol* 2023;**30**:991–1000. https://doi.org/10.1038/s41594-023-01027-2.
39 Malumbres M. Cyclin-dependent kinases. *Genome Biology* 2014;**15**:122. https://doi.org/10.1186/gb4184.
40 Malumbres M, Barbacid M. Mammalian cyclin-dependent kinases. *Trends in Biochemical Sciences* 2005;**30**:630–41. https://doi.org/10.1016/j.tibs.2005.09.005.
41 Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Cancer* 2009;**9**:153–66. https://doi.org/10.1038/nrc2602.
42 Lim S, Kaldis P. Cdks, cyclins and CKIs: roles beyond cell cycle regulation. *Development* 2013;**140**:3079–93. https://doi.org/10.1242/dev.091744.
43 Morgan DO. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu Rev Cell Dev Biol* 1997;**13**:261–91. https://doi.org/10.1146/annurev.cellbio.13.1.261.
44 Pines J. Cyclins and cyclin-dependent kinases: a biochemical view. *Biochemical Journal* 1995;**308**:697–711. https://doi.org/10.1042/bj3080697.
45 Dubbury SJ, Boutz PL, Sharp PA. CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. *Nature* 2018;**564**:141–5. https://doi.org/10.1038/s41586-018-0758-y.
46 Krajewska M, Dries R, Grassetti AV, Dust S, Gao Y, Huang H, *et al.* CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. *Nat Commun* 2019;**10**:1757. https://doi.org/10.1038/s41467-019-09703-y.
47 Fan Z, Devlin JR, Hogg SJ, Doyle MA, Harrison PF, Todorovski I, *et al.* CDK13 cooperates with CDK12 to control global RNA polymerase II processivity. *Science Advances* 2020;**6**:eaaz5041. https://doi.org/10.1126/sciadv.aaz5041.
48 Tellier M, Zaborowska J, Caizzi L, Mohammad E, Velychko T, Schwalb B, *et al.* CDK12 globally stimulates RNA polymerase II transcription elongation and carboxyl-terminal

domain phosphorylation. *Nucleic Acids Research* 2020;**48**:7712–27. https://doi.org/10.1093/nar/gkaa514.
49 Naro C, Bielli P, Sette C. Oncogenic dysregulation of pre-mRNA processing by protein kinases: challenges and therapeutic opportunities. *The FEBS Journal* 2021;**288**:6250–72. https://doi.org/10.1111/febs.16057.
50 Hertel KJ, Graveley BR. RS domains contact the pre-mRNA throughout spliceosome assembly. *Trends in Biochemical Sciences* 2005;**30**:115–8. https://doi.org/10.1016/j.tibs.2005.01.002.
51 Panzeri V, Pieraccioli M, Cesari E, de la Grange P, Sette C. CDK12/13 promote splicing of proximal introns by enhancing the interaction between RNA polymerase II and the splicing factor SF3B1. *Nucleic Acids Research* 2023;**51**:5512–26. https://doi.org/10.1093/nar/gkad258.
52 Knuesel MT, Meyer KD, Bernecky C, Taatjes DJ. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes Dev* 2009;**23**:439–51. https://doi.org/10.1101/gad.1767009.
53 Tsai K-L, Sato S, Tomomori-Sato C, Conaway RC, Conaway JW, Asturias FJ. A conserved Mediator–CDK8 kinase module association regulates Mediator–RNA polymerase II interaction. *Nat Struct Mol Biol* 2013;**20**:611–9. https://doi.org/10.1038/nsmb.2549.
54 Rengachari S, Schilbach S, Aibara S, Dienemann C, Cramer P. Structure of the human Mediator–RNA polymerase II pre-initiation complex. *Nature* 2021;**594**:129–33. https://doi.org/10.1038/s41586-021-03555-7.
55 Abdella R, Talyzina A, Chen S, Inouye CJ, Tjian R, He Y. Structure of the human Mediator-bound transcription preinitiation complex. *Science* 2021;**372**:52–6. https://doi.org/10.1126/science.abg3074.
56 Chen X, Yin X, Li J, Wu Z, Qi Y, Wang X, *et al.* Structures of the human Mediator and Mediator-bound preinitiation complex. *Science* 2021;**372**:eabg0635. https://doi.org/10.1126/science.abg0635.
57 Akoulitchev S, Chuikov S, Reinberg D. TFIIH is negatively regulated by cdk8-containing mediator complexes. *Nature* 2000;**407**:102–6. https://doi.org/10.1038/35024111.
58 Schachter MM, Merrick KA, Larochelle S, Hirschi A, Zhang C, Shokat KM, *et al.* A Cdk7-Cdk4 T-Loop Phosphorylation Cascade Promotes G1 Progression. *Mol Cell* 2013;**50**:250–60. https://doi.org/10.1016/j.molcel.2013.04.003.
59 Guarducci C, Nardone A, Feiglin A, Migliaccio I, Malorni L, Bonechi M, *et al.* Abstract PD7-12: Inhibition of CDK7 overcomes resistance to CDK4/6 inhibitors in hormone receptor positive breast cancer cells. *Cancer Research* 2019;**79**:PD7-12. https://doi.org/10.1158/1538-7445.SABCS18-PD7-12.
60 Sun B, Mason S, Wilson RC, Hazard SE, Wang Y, Fang R, *et al.* Inhibition of the transcriptional kinase CDK7 overcomes therapeutic resistance in HER2-positive breast cancers. *Oncogene* 2020;**39**:50–63. https://doi.org/10.1038/s41388-019-0953-9.
61 Devault A, Martinez AM, Fesquet D, Labbé JC, Morin N, Tassan JP, *et al.* MAT1 ('menage à trois') a new RING finger protein subunit stabilizing cyclin H-cdk7 complexes in starfish and Xenopus CAK. *The EMBO Journal* 1995;**14**:5027–36. https://doi.org/10.1002/j.1460-2075.1995.tb00185.x.

62 Greber BJ, Toso DB, Fang J, Nogales E. The complete structure of the human TFIIH core complex. *eLife* 2019;**8**:e44771. https://doi.org/10.7554/eLife.44771.
63 Abdulrahman W, Iltis I, Radu L, Braun C, Maglott-Roth A, Giraudon C, *et al.* ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities. *Proceedings of the National Academy of Sciences* 2013;**110**:E633–42. https://doi.org/10.1073/pnas.1213981110.
64 Busso D, Keriel A, Sandrock B, Poterszman A, Gileadi O, Egly J-M. Distinct Regions of MAT1 Regulate cdk7 Kinase and TFIIH Transcription Activities*. *Journal of Biological Chemistry* 2000;**275**:22815–23. https://doi.org/10.1074/jbc.M002578200.
65 He Y, Fang J, Taatjes DJ, Nogales E. Structural visualization of key steps in human transcription initiation. *Nature* 2013;**495**:481–6. https://doi.org/10.1038/nature11991.
66 Fisher RP, Morgan DO. A novel cyclin associates with M015/CDK7 to form the CDK-activating kinase. *Cell* 1994;**78**:713–24. https://doi.org/10.1016/0092-8674(94)90535-5.
67 Shiekhattar R, Mermelstein F, Fisher RP, Drapkin R, Dynlacht B, Wessling HC, *et al.* Cdk-activating kinase complex is a component of human transcription factor TFIIH. *Nature* 1995;**374**:283–7. https://doi.org/10.1038/374283a0.
68 Fisher RP. Cdk7: a kinase at the core of transcription and in the crosshairs of cancer drug discovery. *Transcription* 2018;**10**:47–56. https://doi.org/10.1080/21541264.2018.1553483.
69 Ebmeier CC, Erickson B, Allen BL, Allen MA, Kim H, Fong N, *et al.* Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications. *Cell Reports* 2017;**20**:1173–86. https://doi.org/10.1016/j.celrep.2017.07.021.
70 Clopper KC, Taatjes DJ. Chemical inhibitors of transcription-associated kinases. *Current Opinion in Chemical Biology* 2022;**70**:102186. https://doi.org/10.1016/j.cbpa.2022.102186.
71 Larochelle S, Amat R, Glover-Cutter K, Sansó M, Zhang C, Allen JJ, *et al.* Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nat Struct Mol Biol* 2012;**19**:1108–15. https://doi.org/10.1038/nsmb.2399.
72 Hu S, Marineau JJ, Rajagopal N, Hamman KB, Choi YJ, Schmidt DR, *et al.* Discovery and Characterization of SY-1365, a Selective, Covalent Inhibitor of CDK7. *Cancer Research* 2019;**79**:3479–91. https://doi.org/10.1158/0008-5472.CAN-19-0119.
73 Marineau JJ, Hamman KB, Hu S, Alnemy S, Mihalich J, Kabro A, *et al.* Discovery of SY-5609: A Selective, Noncovalent Inhibitor of CDK7. *J Med Chem* 2022;**65**:1458–80. https://doi.org/10.1021/acs.jmedchem.1c01171.
74 Egloff S. CDK9 keeps RNA polymerase II on track. *Cell Mol Life Sci* 2021;**78**:5543–67. https://doi.org/10.1007/s00018-021-03878-8.
75 Bacon CW, D'Orso I. CDK9: a signaling hub for transcriptional control. *Transcription* 2018;**10**:57–75. https://doi.org/10.1080/21541264.2018.1523668.
76 Richters A, Doyle SK, Freeman DB, Lee C, Leifer BS, Jagannathan S, *et al.* Modulating Androgen Receptor-Driven Transcription in Prostate Cancer with Selective CDK9 Inhibitors. *Cell Chemical Biology* 2021;**28**:134-147.e14. https://doi.org/10.1016/j.chembiol.2020.10.001.
77 Jonkers I, Kwak H, Lis JT. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* 2014;**3**:e02407. https://doi.org/10.7554/eLife.02407.

78 Marshall NF, Price DH. Purification of P-TEFb, a Transcription Factor Required for the Transition into Productive Elongation (∗). *Journal of Biological Chemistry* 1995;**270**:12335–8. https://doi.org/10.1074/jbc.270.21.12335.
79 Lis JT, Mason P, Peng J, Price DH, Werner J. P-TEFb kinase recruitment and function at heat shock loci. *Genes Dev* 2000;**14**:792–803. https://doi.org/10.1101/gad.14.7.792.
80 Liang K, Gao X, Gilmore JM, Florens L, Washburn MP, Smith E, *et al.* Characterization of Human Cyclin-Dependent Kinase 12 (CDK12) and CDK13 Complexes in C-Terminal Domain Phosphorylation, Gene Transcription, and RNA Processing. *Molecular and Cellular Biology* 2015;**35**:928–38. https://doi.org/10.1128/MCB.01426-14.
81 Decker T-M, Forné I, Straub T, Elsaman H, Ma G, Shah N, *et al.* Analog-sensitive cell line identifies cellular substrates of CDK9. *Oncotarget* 2019;**10**:6934–43. https://doi.org/10.18632/oncotarget.27334.
82 Tellier M, Zaborowska J, Neve J, Nojima T, Hester S, Fournier M, *et al.* CDK9 and PP2A regulate RNA polymerase II transcription termination and coupled RNA maturation. *EMBO Rep* 2022;**23**:e54520. https://doi.org/10.15252/embr.202154520.
83 Schier AC, Taatjes DJ. Everything at once: cryo-EM yields remarkable insights into human RNA polymerase II transcription. *Nat Struct Mol Biol* 2021;**28**:540–3. https://doi.org/10.1038/s41594-021-00613-6.
84 Kugel JF, Goodrich JA. Promoter escape limits the rate of RNA polymerase II transcription and is enhanced by TFIIE, TFIIH, and ATP on negatively supercoiled DNA. *Proc Natl Acad Sci U S A* 1998;**95**:9232–7.
85 Saecker RM, Record MT, deHaseth PL. Mechanism of Bacterial Transcription Initiation: RNA Polymerase - Promoter Binding, Isomerization to Initiation-Competent Open Complexes, and Initiation of RNA Synthesis. *Journal of Molecular Biology* 2011;**412**:754–71. https://doi.org/10.1016/j.jmb.2011.01.018.
86 Fishburn J, Tomko E, Galburt E, Hahn S. Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. *Proceedings of the National Academy of Sciences* 2015;**112**:3961–6. https://doi.org/10.1073/pnas.1417709112.
87 Grünberg S, Hahn S. Structural insights into transcription initiation by RNA polymerase II. *Trends in Biochemical Sciences* 2013;**38**:603–11. https://doi.org/10.1016/j.tibs.2013.09.002.
88 Bernecky C, Herzog F, Baumeister W, Plitzko JM, Cramer P. Structure of transcribing mammalian RNA polymerase II. *Nature* 2016;**529**:551–4. https://doi.org/10.1038/nature16482.
89 Akoulitchev S, Mäkelä TP, Weinberg RA, Reinberg D. Requirement for TFIIH kinase activity in transcription by RNA polymerase II. *Nature* 1995;**377**:557–60. https://doi.org/10.1038/377557a0.
90 Max T, Søgaard M, Svejstrup JQ. Hyperphosphorylation of the C-terminal Repeat Domain of RNA Polymerase II Facilitates Dissociation of Its Complex with Mediator*. *Journal of Biological Chemistry* 2007;**282**:14113–20. https://doi.org/10.1074/jbc.M701345200.
91 Taatjes DJ, Näär AM, Andel F, Nogales E, Tjian R. Structure, Function, and Activator-Induced Conformations of the CRSP Coactivator. *Science* 2002;**295**:1058–62. https://doi.org/10.1126/science.1065249.

92 Fant C. *Design and Implementation of a Fully Reconstituted Assay to Investigate Mechanisms of Early Human Pol II Transcription*. University of Colorado Boulder; 2019; 2019.
93 Krumm A, Hickey LB, Groudine M. Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. *Genes Dev* 1995;**9**:559–72. https://doi.org/10.1101/gad.9.5.559.
94 Gressel S, Schwalb B, Cramer P. The pause-initiation limit restricts transcription activation in human cells. *Nat Commun* 2019;**10**:3603. https://doi.org/10.1038/s41467-019-11536-8.
95 Min IM, Waterfall JJ, Core LJ, Munroe RJ, Schimenti J, Lis JT. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev* 2011;**25**:742–54. https://doi.org/10.1101/gad.2005511.
96 Gilchrist DA, Santos GD, Fargo DC, Xie B, Gao Y, Li L, *et al.* Pausing of RNA Polymerase II Disrupts DNA-Specified Nucleosome Organization to Enable Precise Gene Regulation. *Cell* 2010;**143**:540–51. https://doi.org/10.1016/j.cell.2010.10.004.
97 Core L, Adelman K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev* 2019;**33**:960–82. https://doi.org/10.1101/gad.325142.119.
98 Adelman K, Kennedy MA, Nechaev S, Gilchrist DA, Muse GW, Chinenov Y, *et al.* Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. *Proc Natl Acad Sci U S A* 2009;**106**:18207–12. https://doi.org/10.1073/pnas.0910177106.
99 Krumm A, Meulia T, Brunvand M, Groudine M. The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region. *Genes Dev* 1992;**6**:2201–13. https://doi.org/10.1101/gad.6.11.2201.
100 Gilchrist DA, Fromm G, dos Santos G, Pham LN, McDaniel IE, Burkholder A, *et al.* Regulating the regulators: the pervasive effects of Pol II pausing on stimulus-responsive gene networks. *Genes Dev* 2012;**26**:933–44. https://doi.org/10.1101/gad.187781.112.
101 Ghosh A, Shuman S, Lima CD. Structural insights to how mammalian capping enzyme reads the CTD code. *Mol Cell* 2011;**43**:299–310. https://doi.org/10.1016/j.molcel.2011.06.001.
102 Nechaev S, Fargo DC, dos Santos G, Liu L, Gao Y, Adelman K. Global Analysis of Short RNAs Reveals Widespread Promoter-Proximal Stalling and Arrest of Pol II in Drosophila. *Science* 2010;**327**:335–8. https://doi.org/10.1126/science.1181421.
103 Rasmussen EB, Lis JT. In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. *Proc Natl Acad Sci U S A* 1993;**90**:7923–7.
104 Vos SM, Farnung L, Urlaub H, Cramer P. Structure of paused transcription complex Pol II–DSIF–NELF. *Nature* 2018;**560**:601–6. https://doi.org/10.1038/s41586-018-0442-2.
105 Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A, Sugimoto S, *et al.* NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress RNA Polymerase II Elongation. *Cell* 1999;**97**:41–51. https://doi.org/10.1016/S0092-8674(00)80713-8.
106 Aoi Y, Smith ER, Shah AP, Rendleman EJ, Marshall SA, Woodfin AR, *et al.* NELF Regulates a Promoter-Proximal Step Distinct from RNA Pol II Pause-Release. *Molecular Cell* 2020;**78**:261-274.e5. https://doi.org/10.1016/j.molcel.2020.02.014.
107 Mandal SS, Chu C, Wada T, Handa H, Shatkin AJ, Reinberg D. Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter

escape by RNA polymerase II. *Proc Natl Acad Sci U S A* 2004;**101**:7572–7. https://doi.org/10.1073/pnas.0401493101.
108 Missra A, Gilmour DS. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proceedings of the National Academy of Sciences* 2010;**107**:11301–6. https://doi.org/10.1073/pnas.1000681107.
109 Wada T, Takagi T, Yamaguchi Y, Watanabe D, Handa H. Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. *EMBO J* 1998;**17**:7395–403. https://doi.org/10.1093/emboj/17.24.7395.
110 Yamada T, Yamaguchi Y, Inukai N, Okamoto S, Mura T, Handa H. P-TEFb-Mediated Phosphorylation of hSpt5 C-Terminal Repeats Is Critical for Processive Transcription Elongation. *Molecular Cell* 2006;**21**:227–37. https://doi.org/10.1016/j.molcel.2005.11.024.
111 Bernecky C, Plitzko JM, Cramer P. Structure of a transcribing RNA polymerase II–DSIF complex reveals a multidentate DNA–RNA clamp. *Nat Struct Mol Biol* 2017;**24**:809–15. https://doi.org/10.1038/nsmb.3465.
112 Decker T-M. Mechanisms of Transcription Elongation Factor DSIF (Spt4–Spt5). *Journal of Molecular Biology* 2021;**433**:166657. https://doi.org/10.1016/j.jmb.2020.09.016.
113 Cheng B, Price DH. Properties of RNA Polymerase II Elongation Complexes Before and After the P-TEFb-mediated Transition into Productive Elongation*. *Journal of Biological Chemistry* 2007;**282**:21901–12. https://doi.org/10.1074/jbc.M702936200.
114 Fant C, Levandowski C, Gupta K, Maas Z, Moir J, Rubin J, *et al.* TFIID enables RNA polymerase II promoter-proximal pausing. *Mol Cell* 2020;**78**:785-793.e8. https://doi.org/10.1016/j.molcel.2020.03.008.
115 Ghosh SKB, Missra A, Gilmour DS. Negative Elongation Factor Accelerates the Rate at Which Heat Shock Genes Are Shut off by Facilitating Dissociation of Heat Shock Factor ▿. *Mol Cell Biol* 2011;**31**:4232–43. https://doi.org/10.1128/MCB.05930-11.
116 Sansó M, Levin RS, Lipp JJ, Wang VY-F, Greifenberg AK, Quezada EM, *et al.* P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. *Genes Dev* 2016;**30**:117–31. https://doi.org/10.1101/gad.269589.115.
117 Rimel JK, Poss ZC, Erickson B, Maas ZL, Ebmeier CC, Johnson JL, *et al.* Selective inhibition of CDK7 reveals high-confidence targets and new models for TFIIH function in transcription. *Genes Dev* 2020. https://doi.org/10.1101/gad.341545.120.
118 Laitem C, Zaborowska J, Isa NF, Kufs J, Dienstbier M, Murphy S. CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II–transcribed genes. *Nat Struct Mol Biol* 2015;**22**:396–403. https://doi.org/10.1038/nsmb.3000.
119 Kotake Y, Sagane K, Owa T, Mimori-Kiyosue Y, Shimizu H, Uesugi M, *et al.* Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nat Chem Biol* 2007;**3**:570–5. https://doi.org/10.1038/nchembio.2007.16.
120 Schwartz BE, Larochelle S, Suter B, Lis JT. Cdk7 Is Required for Full Activation of Drosophila Heat Shock Genes and RNA Polymerase II Phosphorylation In Vivo. *Mol Cell Biol* 2003;**23**:6876–86. https://doi.org/10.1128/MCB.23.19.6876-6886.2003.

121 Yost HJ, Lindquist S. RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. *Cell* 1986;**45**:185–93. https://doi.org/10.1016/0092-8674(86)90382-X.
122 Liu W, Ma Q, Wong K, Li W, Ohgi K, Zhang J, *et al.* Brd4 and JMJD6-Associated Anti-Pause Enhancers in Regulation of Transcriptional Pause Release. *Cell* 2013;**155**:1581–95. https://doi.org/10.1016/j.cell.2013.10.056.
123 Dawson MA, Prinjha RK, Dittmann A, Giotopoulos G, Bantscheff M, Chan W-I, *et al.* Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature* 2011;**478**:529–33. https://doi.org/10.1038/nature10509.
124 Fitz J, Neumann T, Pavri R. Regulation of RNA polymerase II processivity by Spt5 is restricted to a narrow window during elongation. *The EMBO Journal* 2018;**37**:e97965. https://doi.org/10.15252/embj.201797965.
125 Hou L, Wang Y, Liu Y, Zhang N, Shamovsky I, Nudler E, *et al.* Paf1C regulates RNA polymerase II progression by modulating elongation rate. *Proceedings of the National Academy of Sciences* 2019;**116**:14583–92. https://doi.org/10.1073/pnas.1904324116.
126 Parua PK, Kalan S, Benjamin B, Sansó M, Fisher RP. Distinct Cdk9-phosphatase switches act at the beginning and end of elongation by RNA polymerase II 2020:2020.06.14.150847. https://doi.org/10.1101/2020.06.14.150847.
127 Brannan K, Kim H, Erickson B, Glover-Cutter K, Kim S, Fong N, *et al.* mRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of RNA Polymerase II Transcription. *Molecular Cell* 2012;**46**:311–24. https://doi.org/10.1016/j.molcel.2012.03.006.
128 Wang Y, Zhang T, Kwiatkowski N, Abraham BJ, Lee TI, Xie S, *et al.* CDK7-Dependent Transcriptional Addiction in Triple-Negative Breast Cancer. *Cell* 2015;**163**:174–86. https://doi.org/10.1016/j.cell.2015.08.063.
129 Kwiatkowski N, Zhang T, Rahl PB, Abraham BJ, Reddy J, Ficarro SB, *et al.* Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* 2014;**511**:616–20. https://doi.org/10.1038/nature13393.
130 Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, *et al.* Remodeling of the Enhancer Landscape during Macrophage Activation Is Coupled to Enhancer Transcription. *Molecular Cell* 2013;**51**:310–25. https://doi.org/10.1016/j.molcel.2013.07.010.
131 Robinson PJ, Trnka MJ, Bushnell DA, Davis R, Mattei P-J, Burlingame AL, *et al.* Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* 2016;**166**:1411-1422.e16. https://doi.org/10.1016/j.cell.2016.08.050.
132 Myers LC, Kornberg RD. Mediator of transcriptional regulation. *Annu Rev Biochem* 2000;**69**:729–49. https://doi.org/10.1146/annurev.biochem.69.1.729.
133 Komarnitsky P, Cho E-J, Buratowski S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* 2000;**14**:2452–60.
134 Schroeder SC, Schwer B, Shuman S, Bentley D. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev* 2000;**14**:2435–40.
135 Zaborowska J, Egloff S, Murphy S. The pol II CTD: new twists in the tail. *Nat Struct Mol Biol* 2016;**23**:771–7. https://doi.org/10.1038/nsmb.3285.

136 Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, *et al.* c-Myc Regulates Transcriptional Pause Release. *Cell* 2010;**141**:432–45. https://doi.org/10.1016/j.cell.2010.03.030.
137 Kinyamu HK, Bennett BD, Bushel PR, Archer TK. Proteasome inhibition creates a chromatin landscape favorable to RNA Pol II processivity. *J Biol Chem* 2020;**295**:1271–87. https://doi.org/10.1074/jbc.RA119.011174.
138 Schwartz JC, Ebmeier CC, Podell ER, Heimiller J, Taatjes DJ, Cech TR. FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. *Genes Dev* 2012;**26**:2690–5. https://doi.org/10.1101/gad.204602.112.
139 Vihervaara A, Versluis P, Lis J. *PRO-IP-seq Tracks Molecular Modifications of Engaged Pol II Complexes at Nucleotide Resolution*. Genomics; 2023.
140 Czudnochowski N, Bösken CA, Geyer M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat Commun* 2012;**3**:842. https://doi.org/10.1038/ncomms1846.
141 Tietjen JR, Zhang DW, Rodríguez-Molina JB, White BE, Akhtar MS, Heidemann M, *et al.* Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol* 2010;**17**:1154–61. https://doi.org/10.1038/nsmb.1900.
142 Yankulov K, Yamashita K, Roy R, Egly J-M, Bentley DL. The Transcriptional Elongation Inhibitor 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole Inhibits Transcription Factor IIH-associated Protein Kinase (∗). *Journal of Biological Chemistry* 1995;**270**:23922–5. https://doi.org/10.1074/jbc.270.41.23922.
143 Senderowicz AM. Flavopiridol: the first cyclin-dependent kinase inhibitor in human clinical trials. *Invest New Drugs* 1999;**17**:313–20. https://doi.org/10.1023/a:1006353008903.
144 Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet* 2020;**21**:597–614. https://doi.org/10.1038/s41576-020-0236-x.
145 Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, *et al.* Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun* 2019;**10**:1784. https://doi.org/10.1038/s41467-018-08148-z.
146 Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, *et al.* A global reference for human genetic variation. *Nature* 2015;**526**:68–74. https://doi.org/10.1038/nature15393.
147 Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol* 2017;**35**:316–9. https://doi.org/10.1038/nbt.3820.
148 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841–2. https://doi.org/10.1093/bioinformatics/btq033.
149 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* 2021;**10**:giab008. https://doi.org/10.1093/gigascience/giab008.
150 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019;**37**:907–15. https://doi.org/10.1038/s41587-019-0201-4.
151 Bushnell B. BBMap: A Fast, Accurate, Splice-Aware Aligner 2014.

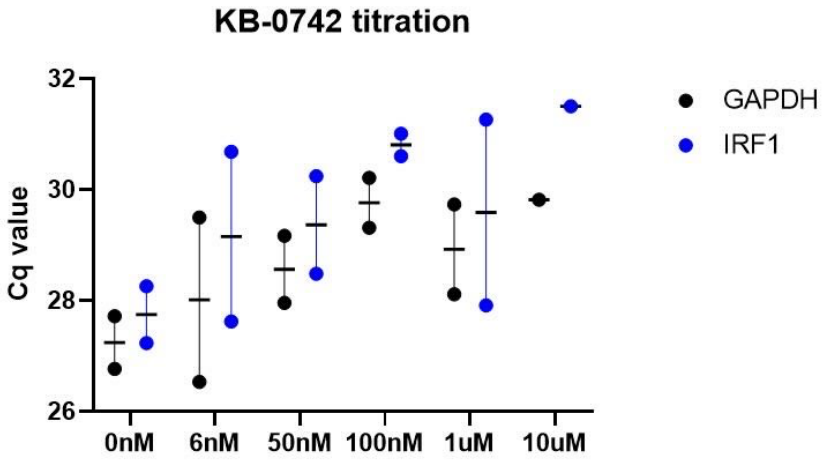
152 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;**29**:24–6. https://doi.org/10.1038/nbt.1754.
153 Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 2012;**28**:2184–5. https://doi.org/10.1093/bioinformatics/bts356.
154 *The Python Language Reference*. Python Documentation. n.d. URL: https://docs.python.org/3/reference/index.html (Accessed 2 October 2023).
155 Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016;**32**:3047–8. https://doi.org/10.1093/bioinformatics/btw354.
156 Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 2014;**42**:W187–91. https://doi.org/10.1093/nar/gku365.
157 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetJournal* 2011;**17**:10–2. https://doi.org/10.14806/ej.17.1.200.
158 Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* 2019;**9**:9354. https://doi.org/10.1038/s41598-019-45839-z.
159 Morales J, Pujar S, Loveland JE, Astashyn A, Bennett R, Berry A, *et al.* A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature* 2022;**604**:310–5. https://doi.org/10.1038/s41586-022-04558-8.
160 Der SD, Zhou A, Williams BRG, Silverman RH. Identification of genes differentially regulated by interferon α, β, or γ using oligonucleotide arrays. *Proceedings of the National Academy of Sciences* 1998;**95**:15623–8. https://doi.org/10.1073/pnas.95.26.15623.
161 Steinparzer I, Sedlyarov V, Rubin JD, Eislmayr K, Galbraith MD, Levandowski CB, *et al.* Transcriptional Responses to IFN-γ Require Mediator Kinase-Dependent Pause Release and Mechanistically Distinct CDK8 and CDK19 Functions. *Molecular Cell* 2019;**76**:485-499.e8. https://doi.org/10.1016/j.molcel.2019.07.034.
162 Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;**483**:603–7. https://doi.org/10.1038/nature11003.
163 Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, *et al.* The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* 2016;**2016**:baw100. https://doi.org/10.1093/database/baw100.
164 Mayer A, Lidschreiber M, Siebert M, Leike K, Söding J, Cramer P. Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* 2010;**17**:1272–8. https://doi.org/10.1038/nsmb.1903.
165 Cheng C, Sharp PA. RNA Polymerase II Accumulation in the Promoter-Proximal Region of the Dihydrofolate Reductase and γ-Actin Genes. *Mol Cell Biol* 2003;**23**:1961–7. https://doi.org/10.1128/MCB.23.6.1961-1967.2003.
166 Glover-Cutter K, Larochelle S, Erickson B, Zhang C, Shokat K, Fisher RP, *et al.* TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II.

Molecular and Cellular Biology 2009;**29**:5455–64. https://doi.org/10.1128/MCB.00637-09.

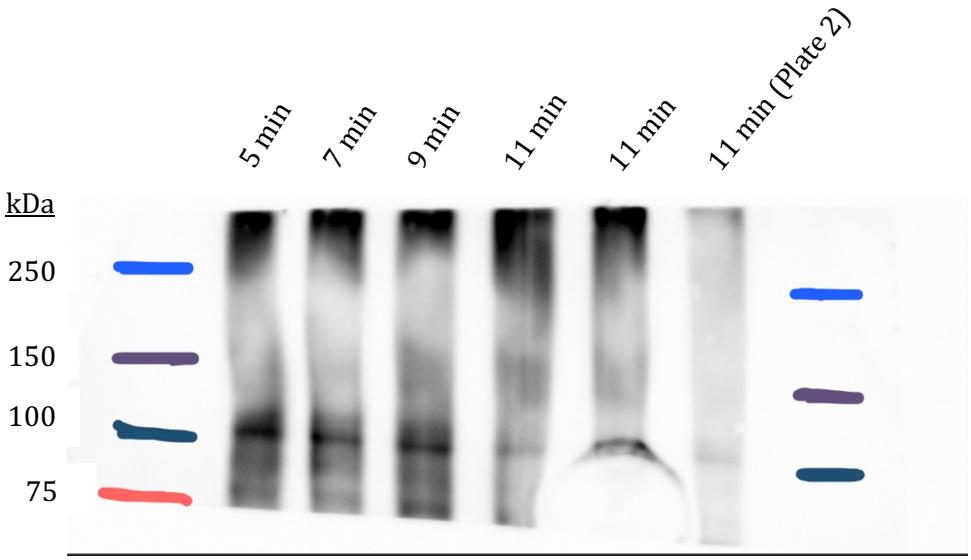
167 Kim H, Erickson B, Luo W, Seward D, Graber JH, Pollock DD, *et al.* Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* 2010;**17**:1279–86. https://doi.org/10.1038/nsmb.1913.

168 Bataille AR, Jeronimo C, Jacques P-É, Laramée L, Fortin M-È, Forest A, *et al.* A Universal RNA Polymerase II CTD Cycle Is Orchestrated by Complex Interplays between Kinase, Phosphatase, and Isomerase Enzymes along Genes. *Molecular Cell* 2012;**45**:158–70. https://doi.org/10.1016/j.molcel.2011.11.024.

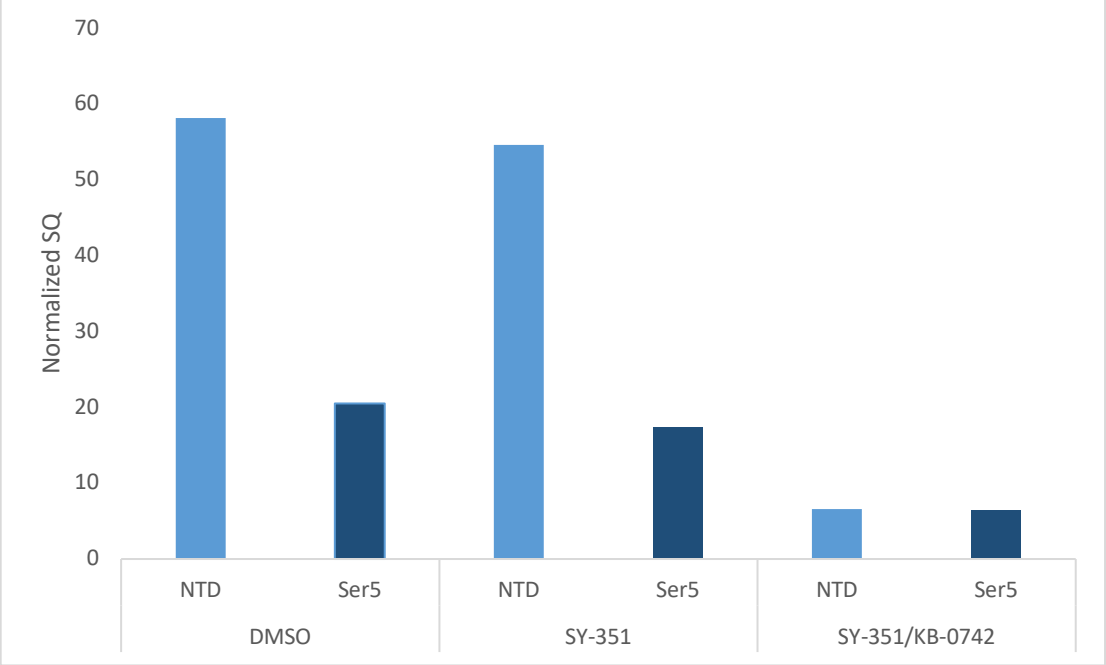
6. APPENDIX



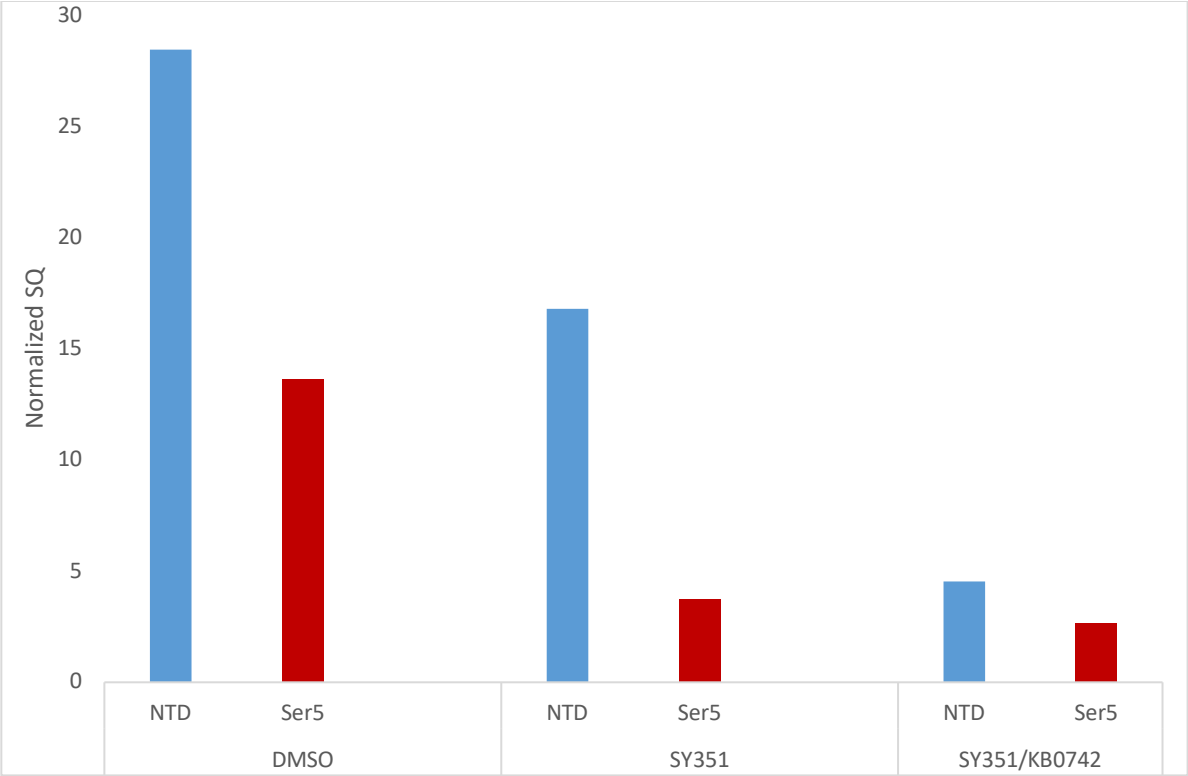
S1: qPCR Titration of KB-0742 to Assess Ideal Assay Concentrations. DNA from HCT116 cells titrated with KB-0742 was isolated, and relative amounts of expression at IRF1 and GAPDH were quantified. Despite a large amount of error in the qPCR, the data indicated that 100 nM KB-0742 was optimal. The IC_{50} of KB-0742 is 6 nM[76]. Data collection and processing performed with the assistance of Olivia Luyties.



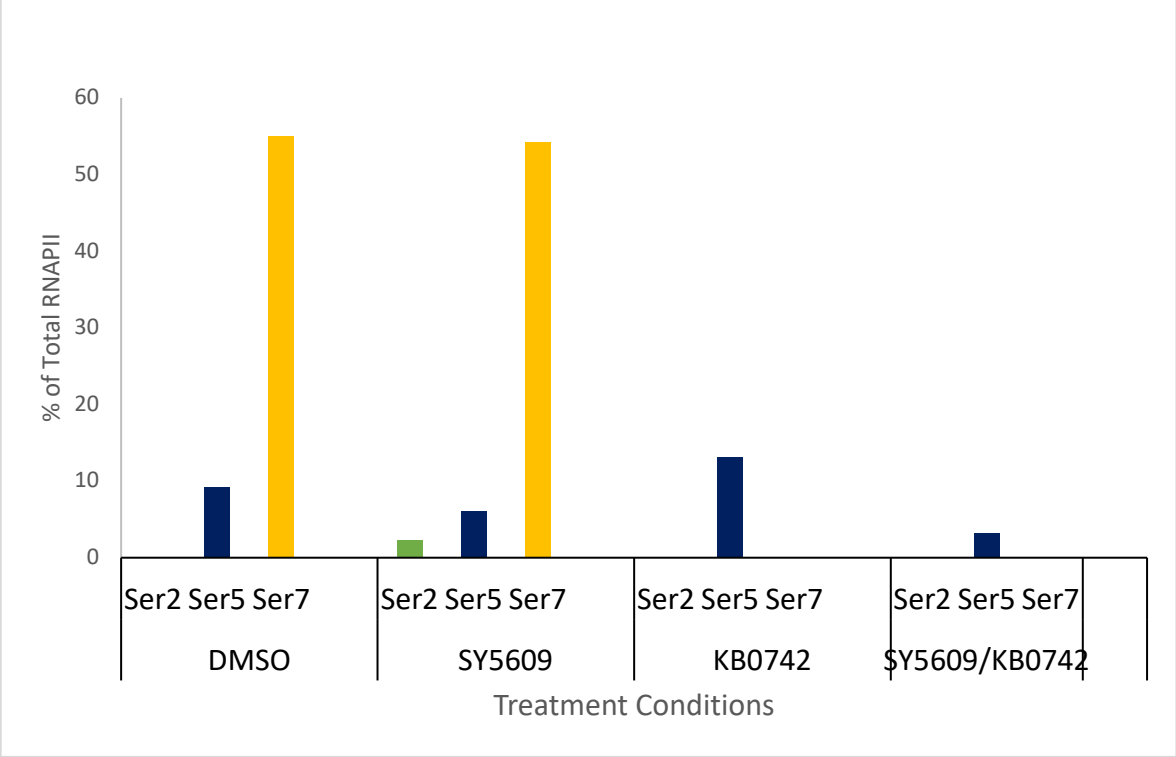
***S2*: Western blot for RNAPII NTD to Assess Appropriate Shearing Timepoints.** Two 15 cm plates of HCT116 cells were grown to 70-90% confluency (roughly 25 million cells total) and crosslinked and sheared as described in the methods section. At each timepoint, 10 uL of sheared chromatin was aliquoted and frozen at -20 C. All samples shown above originated from Plate 1 unless otherwise noted. Samples were then thawed on ice and mixed with 3 uL SDS-PAGE loading buffer. 6 uL of Bio-Rad Kaleidescope ladder (CAT # 1610375) was used. Samples were ran on a Bio-Rad MiniProtean TGX 4-20% Gradient (CAT#4561096) gel at 165 V until the loading dye front ran off the gel. Proteins were transferred to a nitrocellulose membrane for 45 minutes at 100 V in transfer buffer (1 part methanol, 1 part 10x running buffer, 8 parts cold MilliQ water) on ice. The membrane was incubated for ten minutes in milk blocker (5 g powdered milk/100 mL 1x TBST buffer), followed by incubation in milk blocker plus a 1:1000 R pb1 NTD D8L4Y Rabbit mAb: milk blocker dilution overnight at 4°C. Antibody was discarded and the membrane was washed 3x5 minutes with TBST buffer. Membrane was incubated with anti-Rabbit secondary antibody at a 1:5000 dilution in milk blocker for 1 hour. Secondary antibody was discarded and membrane was washed 3x5 minutes with TBST buffer. Membrane was incubated with Millipore Immobilon Western Chemiluminescent HRP Substrate (CAT #WBKL50100) imaged on the LAS Imaging System on the chemiluminescent setting with exposure time set to automatic.



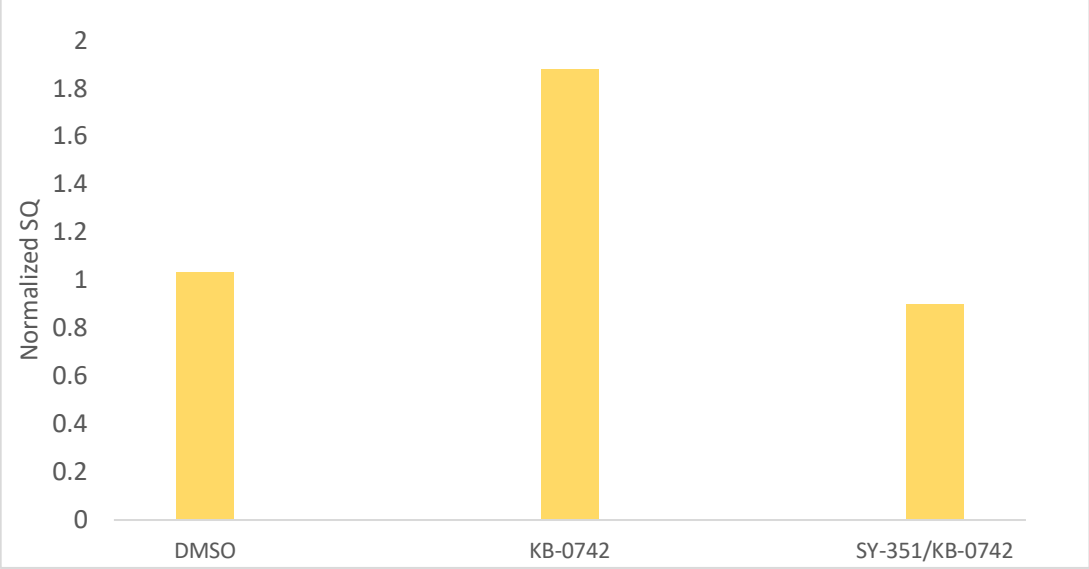
S3: qPCR Data from SY-351 treated samples at the *IRF1* promoter region locus. 0.5 ng of NTD samples and 0.1 ng of pSer5 samples were used in each reaction. Input samples from each IP were included at the same amount. All reactions were performed in triplicate. Briefly, the aforementioned amounts of each sample were mixed with 400 uM each of forward and reverse primer for the region just upstream of the *IRF1* TSS (Table 6). Sample and primer were mixed with 2x SYBR Green PowerUp Mastermix, and aliquoted into 10 uL reactions. A standard curve was generated using DNA from SY-351 treated pSer5 input and included a no DNA control. The standard curve (R^2=0.9971, PCR efficiency = 83%) was used to generate starting quantities (SQ) from all samples. Normalized SQ values were calculated by dividing the SQ of a sample by that of its corresponding input.



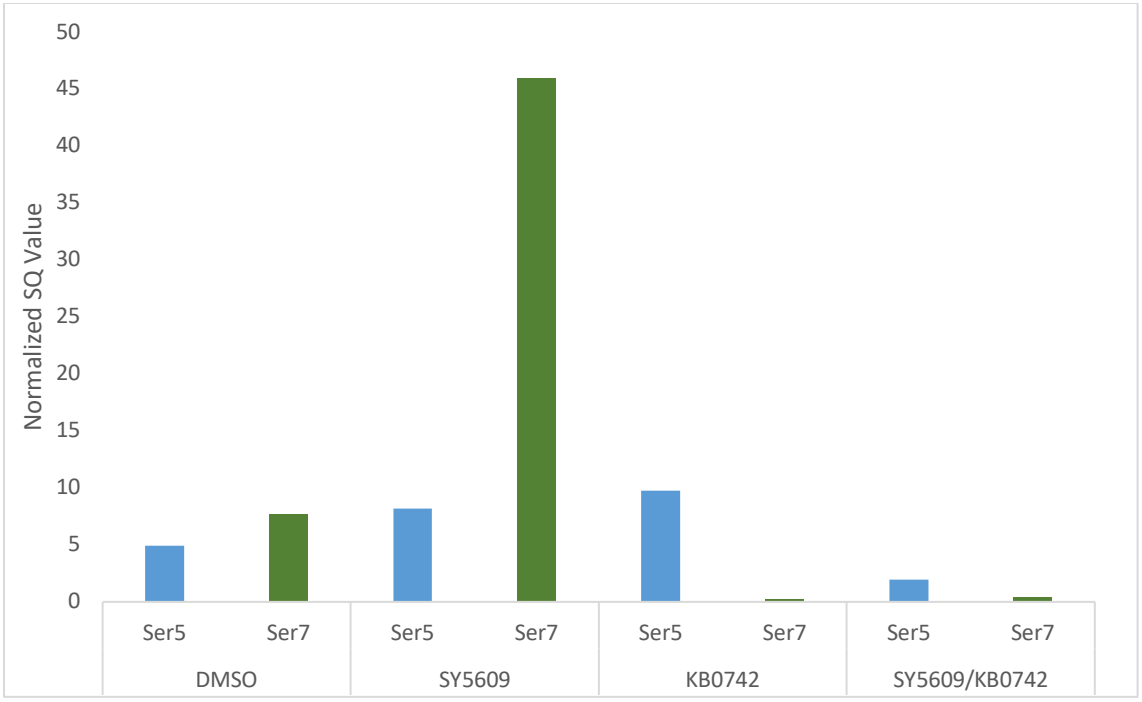
S4: qPCR Data from SY-351 treated samples at the *B2M* promoter region. 0.4 ng of NTD samples and 0.1 ng of pSer5 samples were used in each reaction. Input samples from each IP were included at the same amount. All reactions were performed in triplicate. Briefly, the aforementioned amounts of each sample were mixed with 400 uM each of forward and reverse primer for the region just upstream of the *B2M* TSS (Table 6). Sample and primer were mixed with 2x SYBR Green PowerUp Mastermix, and aliquoted into 10 uL reactions. A standard curve was generated using DNA from SY/KB treated pSer5 input and included a no DNA control. The standard curve (R^2=0.9984, PCR efficiency = 93%) was used to generate starting quantities (SQ) from all samples. Normalized SQ values were calculated by dividing the SQ of a sample by that of its corresponding input.



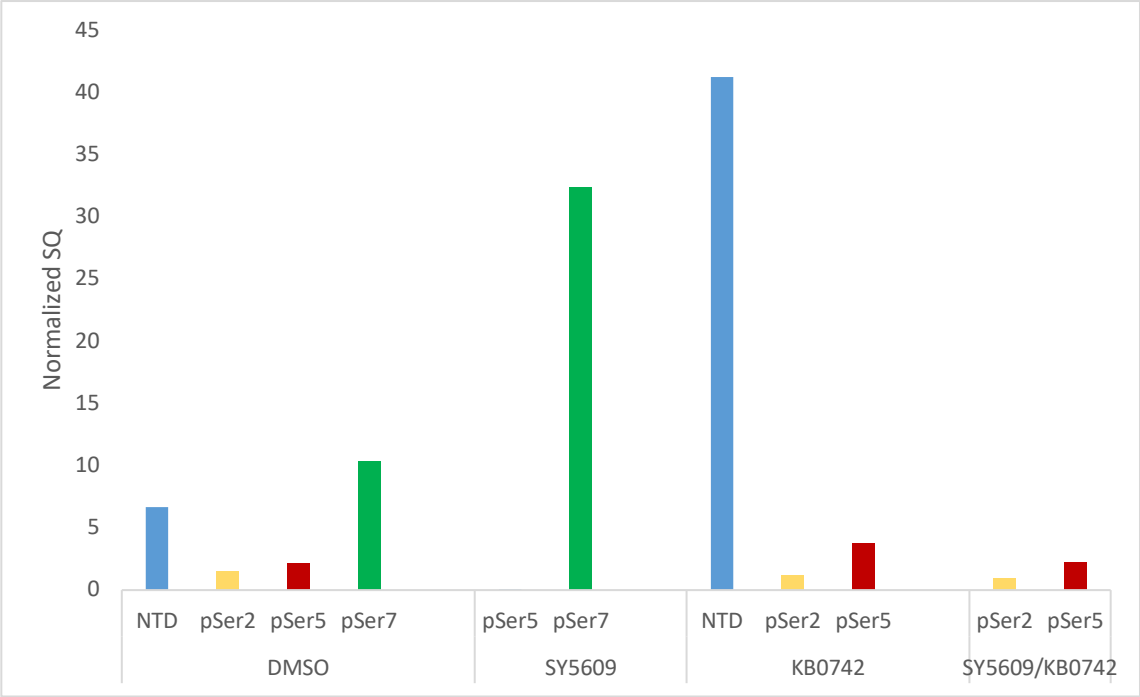
S5: qPCR Data from SY-5609 B1 samples at the *IRF1* promoter region. 0.4 ng of NTD samples, 0.1 ng of pSer5/7 samples, and 0.3 ng of pSer2 samples were used in each reaction. Input samples from each IP were included at the same amount. All reactions were performed in triplicate. Briefly, the aforementioned amounts of each sample were mixed with 400 uM each of forward and reverse primer for the region just upstream of the *IRF1* TSS (Table 6). Sample and primer were mixed with Applied Biosystems PowerUp SYBR Green Mastermix (CAT #A25742), and aliquoted into 10 uL reactions. A standard curve was generated using DNA from SY-5609 treated pSer5 input and included a no DNA control. The standard curve (R^2=0.9962, PCR efficiency = 86%) was used to generate starting quantities (SQ) from all samples. Normalized SQ values were calculated by dividing the SQ of a sample by that of its corresponding input. SQ's of phosphomarks were then divided by those of NTD samples to calculate a phosphomarks as a % of total protein. Serine 2 samples fell below the limit of detection in all samples but the SY5609 treated group. Serine 7 fell below the limit of detection in the KB-0742 and SY/KB samples.



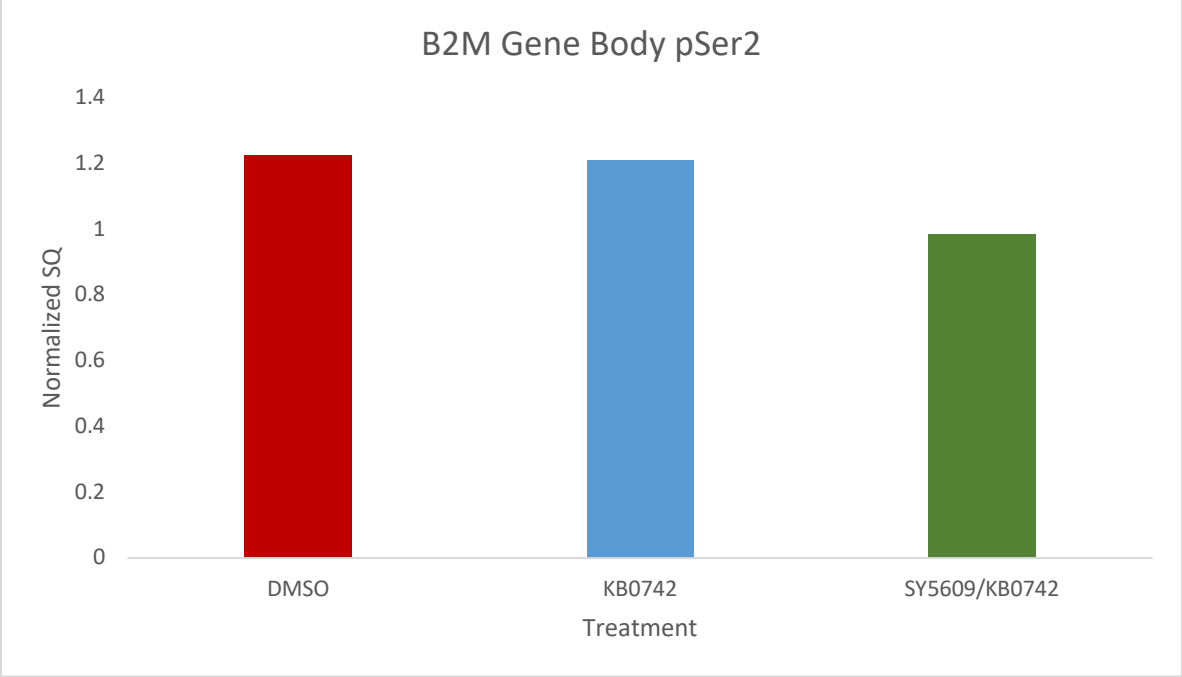
S6: qPCR Data from SY5609 B1 pSer2 samples at the *IRF1* gene body. 0.3 ng of pSer2 samples were used in each reaction. Input samples from each IP were included at the same amount. All reactions were performed in triplicate. SY-5609 treated pSer2 DNA was not included due to low yield. Briefly, the aforementioned amounts of each sample were mixed with 400 uM each of forward and reverse primer for the gene body region ~1000 bp downstream of the *IRF1* TSS (Table 6). Sample and primer were mixed with 2x SYBR Green PowerUp Mastermix, and aliquoted into 10 uL reactions. A standard curve was generated using DNA from SY/KB treated pSer2 input and included a no DNA control. The standard curve (R^2=0.979, PCR efficiency = 89%) was used to generate starting quantities (SQ) from all samples. Normalized SQ values were calculated by dividing the SQ of a sample by that of its corresponding input.



S7: qPCR Data from SY5609 B1 pSer5 and pSer7 samples at the *IRF1* promoter region. 0.1 ng of pSer5/7 samples were used in each reaction. Input samples from each IP were included at the same amount. All reactions were performed in triplicate. Briefly, the aforementioned amounts of each sample were mixed with 400 uM each of forward and reverse primer for the region just upstream of the *IRF1* TSS (Table 6). Sample and primer were mixed with 2x SYBR Green PowerUp Mastermix, and aliquoted into 10 uL reactions. A standard curve was generated using DNA from KB-0742 treated pSer5 input and included a no DNA control. The standard curve (R^2=0.9924, PCR efficiency = 82%) was used to generate starting quantities (SQ) from all samples. Normalized SQ values were calculated by dividing the SQ of a sample by that of its corresponding input. Serine 2 samples fell below the limit of detection in all samples but the SY5609 treated group. Serine 7 fell below the limit of detection in the KB-0742 and SY/KB samples.



S8: qPCR Data from SY-5609 B1 samples at the *B2M* promoter region. 0.4 ng of NTD samples, 0.1 ng of pSer5/7 samples, and 0.3 ng pSer2 samples were used in each reaction. Input samples from each IP were included at the same amount. All reactions were performed in triplicate. Samples not shown were not included due to yield limitations. Briefly, the aforementioned amounts of each sample were mixed with 400 uM each of forward and reverse primer for the region just upstream of the *B2M* TSS (Table 6). Sample and primer were mixed with 2x SYBR Green PowerUp Mastermix, and aliquoted into 10 uL reactions. A standard curve was generated using DNA from SY/KB treated pSer5 input and included a no DNA control. The standard curve (R^2=0.9660, PCR efficiency = 90%) was used to generate starting quantities (SQ) from all samples. Normalized SQ values were calculated by dividing the SQ of a sample by that of its corresponding input.



S9: qPCR Data from SY5609 B1 pSer2 samples at the *B2M* gene body. 0.3 ng of pSer2 samples were used in each reaction. Input samples from each IP were included at the same amount. All reactions were performed in triplicate. SY-5609 treated pSer2 DNA was not included due to low yield. Briefly, the aforementioned amounts of each sample were mixed with 400 uM each of forward and reverse primer for the gene body region ~700 bp downstream of the *B2M* TSS (Table 6). Sample and primer were mixed with 2x SYBR Green PowerUp Mastermix, and aliquoted into 10 uL reactions. A standard curve was generated using DNA from SY-5609 treated pSer2 input and included a no DNA control. The standard curve (R^2=0.9955, PCR efficiency = 99%) was used to generate starting quantities (SQ) from all samples. Normalized SQ values were calculated by dividing the SQ of a sample by that of its corresponding input.

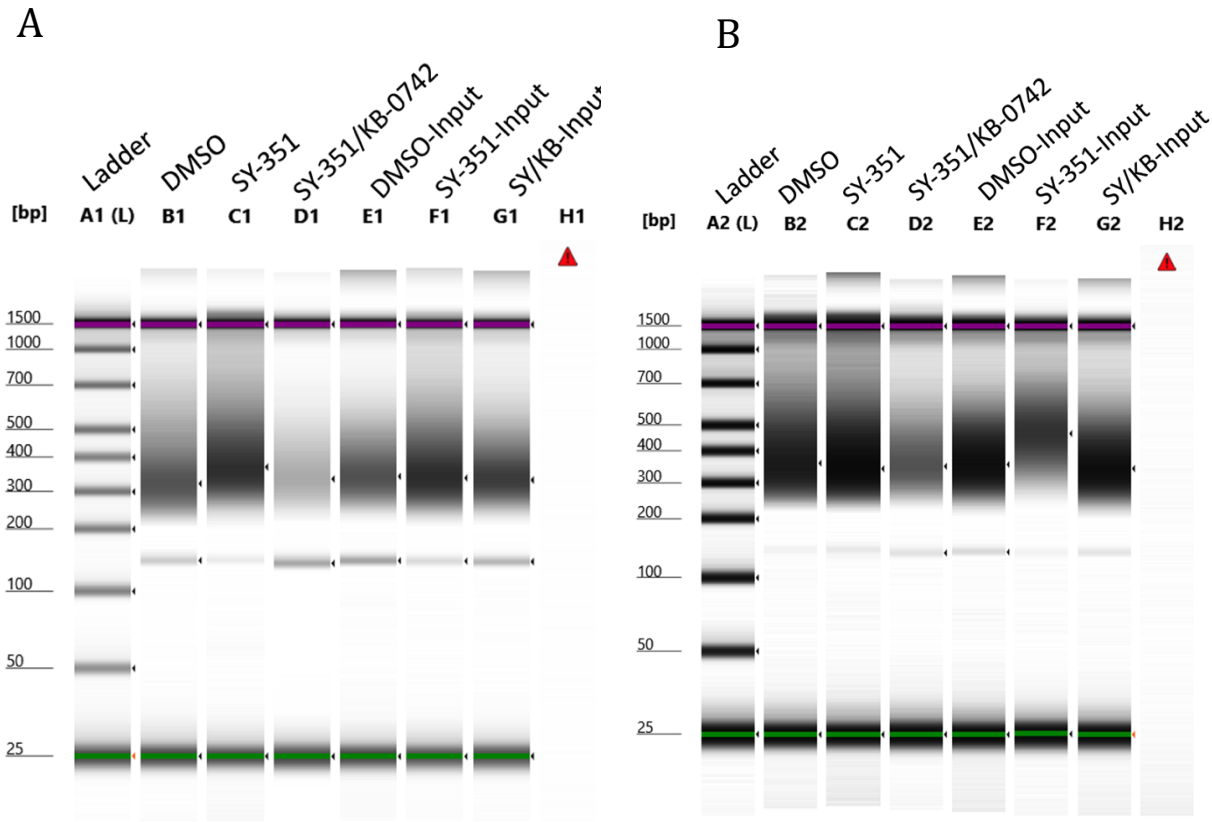
Treatment	IP	Technical Replicate	Total DNA (ng)
DMSO	NTD	1	26.00
DMSO	NTD	2	20.60
DMSO	pSer5	1	3.60
DMSO	pSer5	2	7.44
SY-351	NTD	1	10.76
SY-351	NTD	2	14.52
SY-351	pSer5	1	2.48
SY-351	pSer5	2	10.24
SY-351/KB-0742	NTD	1	70.8
SY-351/KB-0742	NTD	2	83.2
SY-351/KB-0742	pSer5	1	2.68
SY-351/KB-0742	pSer5	2	12.32

S10: Table of total DNA yields from ChIP Immunoprecipitations of SY-351 Experiment Group. All samples were resuspended in 20 uL 0.1x TE Buffer. Input samples (not shown) were resuspended in 30 uL 0.1x TE Buffer.

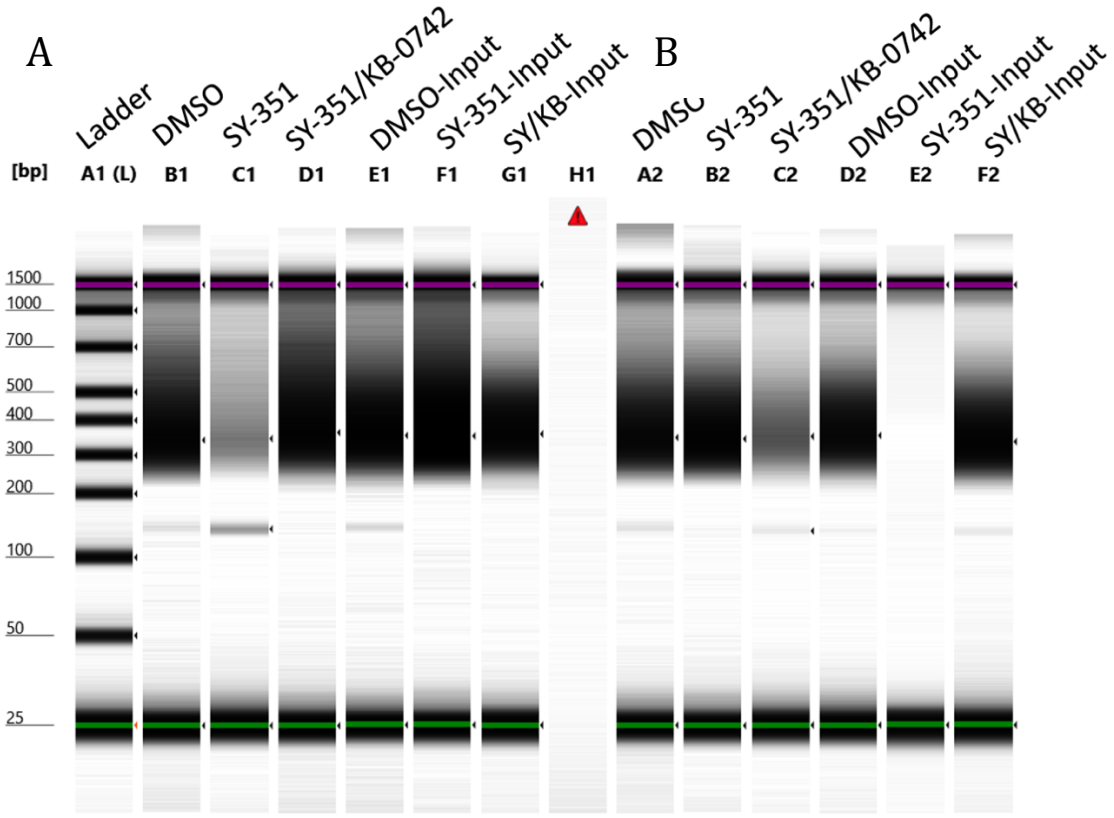
Treatment	IP	Biological Replicate	Total DNA (ng)
DMSO	NTD	1	125.2
DMSO	NTD	2	528.2
DMSO	pSer2	1	52.8
DMSO	pSer2	2	33.6
DMSO	pSer5	1	53.6
DMSO	pSer5	2	44.08
DMSO	pSer7	1	56.8
DMSO	pSer7	2	13.19
SY-5609	NTD	1	9.96
SY-5609	NTD	2	37.43
SY-5609	pSer2	1	5.4
SY-5609	pSer2	2	8.56
SY-5609	pSer5	1	16.44
SY-5609	pSer5	2	42.56
SY-5609	pSer7	1	<0.05
SY-5609	pSer7	2	16.07
KB-0742	NTD	1	27.8
KB-0742	NTD	2	10.98
KB-0742	pSer2	1	12.48
KB-0742	pSer2	2	9.64
KB-0742	pSer5	1	42.8
KB-0742	pSer5	2	6.12
KB-0742	pSer7	1	5.08
KB-0742	pSer7	2	5.93

SY-5609/KB-0742	NTD	1	6.76
SY-5609/KB-0742	NTD	2	6.73
SY-5609/KB-0742	pSer2	1	35
SY-5609/KB-0742	pSer2	2	2.88
SY-5609/KB-0742	pSer5	1	47.2
SY-5609/KB-0742	pSer5	2	3.12
SY-5609/KB-0742	pSer7	1	3.32
SY-5609/KB-0742	pSer7	2	6.35
DMSO	No Ab	2	3.96

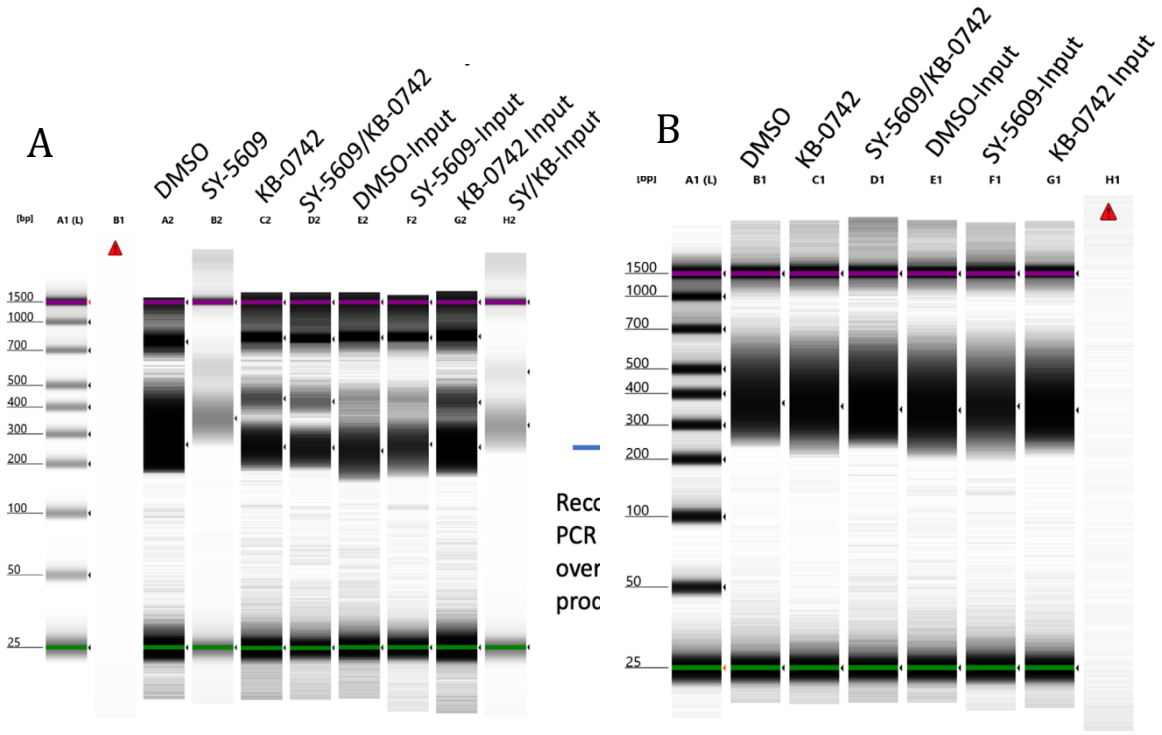
S11: Table of total DNA yields from ChIP Immunoprecipitations of SY-5609 Experiment Groups. All DNA was resuspended in 20 uL of 0.1x TE Buffer. Input samples (not shown) were resuspended in 30 uL 0.1x TE Buffer.



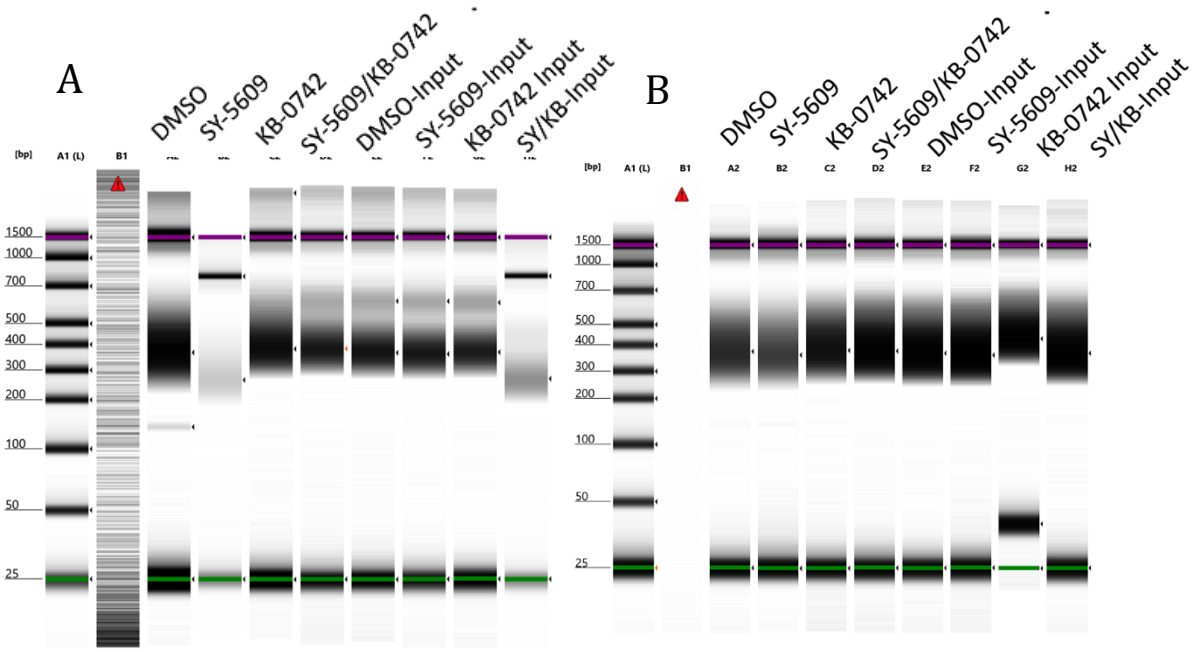
S12: SY-351 RNAPII NTD T1 (A) & T2 (B) Libraries Visualized on an Agilent 2200 Tapestation. 1 µL of each sample or D1000 Ladder was mixed with 3 µL Agilent D1000 Tapestation Buffer. Gels are imaged at roughly 60% saturation. All libraries were amplified for 16 PCR cycles.



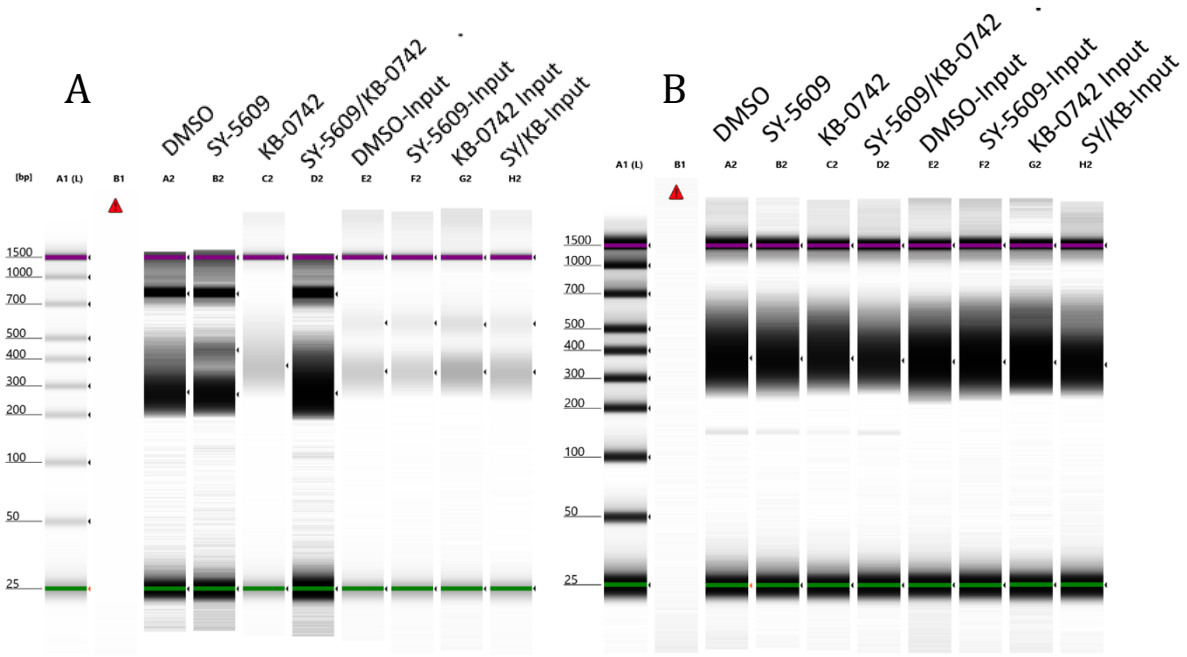
S13: SY351 S5 T1 (A) and T2 (B) Libraries Visualized on an Agilent 2200 Tapestation. 1 µL of each sample or D1000 Ladder was mixed with 3 µL Agilent D1000 Tapestation Buffer. Gels are imaged at roughly 60% saturation. All libraries were amplified for 16 cycles.



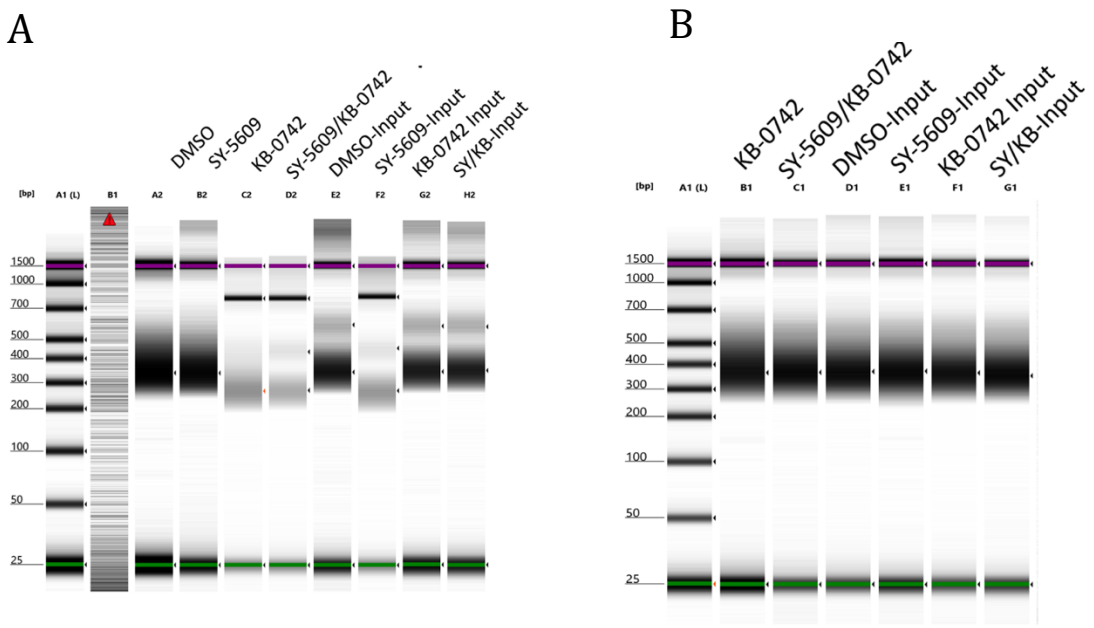
S14: 5609 B1 NTD Libraries Visualized on an Agilent 2200 Tapestation. 1 µL of each sample or D1000 Ladder was mixed with 3 µL Agilent D1000 Tapestation Buffer. Gels are imaged at roughly 90% and 60% saturation respectively. A) All libraries were amplified for 16 cycles. B) Most libraries subjected to reconditioning PCR display anticipated size distribution (~350 bp). Reconditioning was performed by taking 16 uL of overamplified library and subjecting it to round of PCR amplification, followed by a 0.8x bead based cleanup.



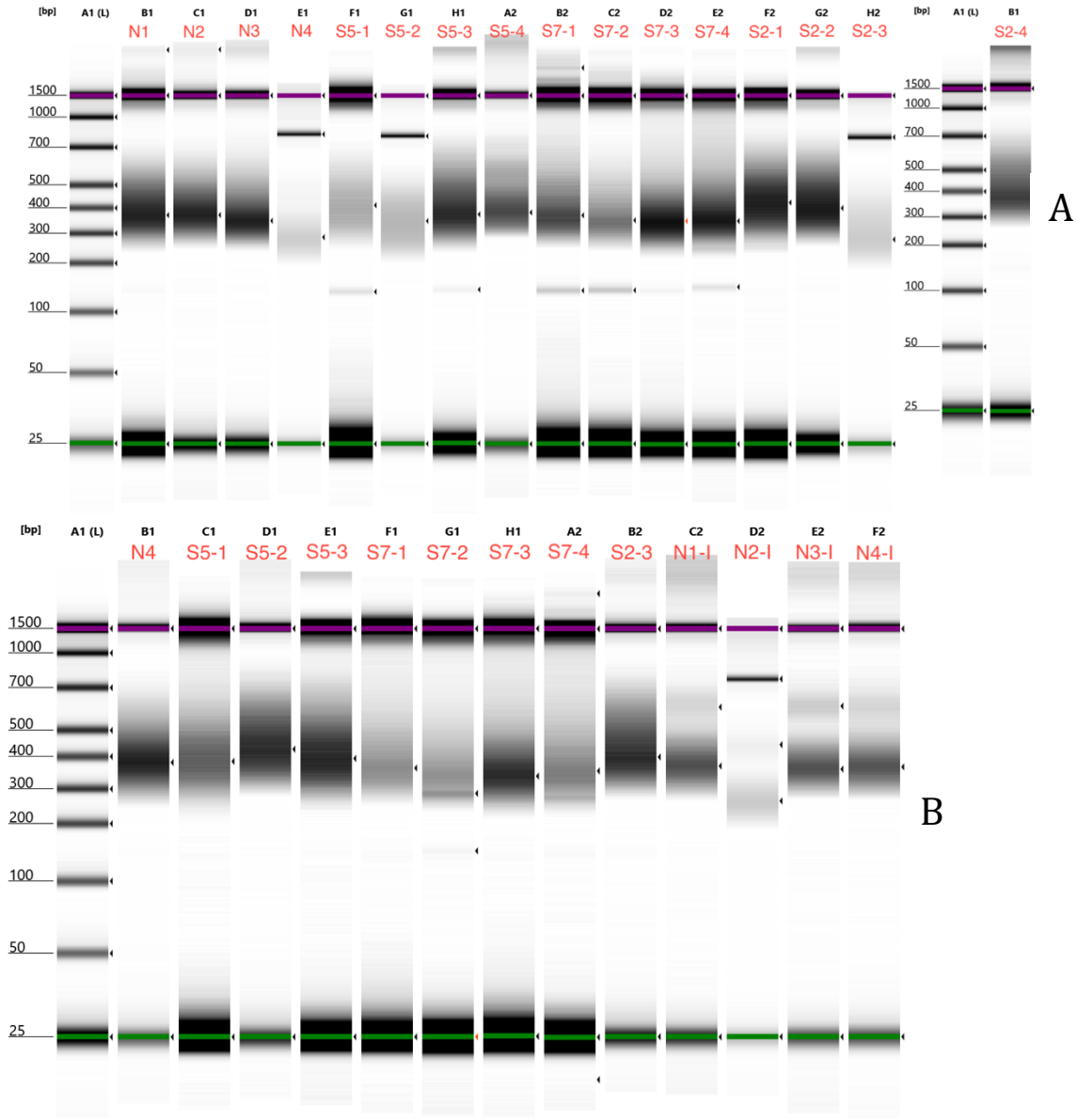
S15: 5609 B1 pSer5 Libraries Visualized on an Agilent 2200 Tapestation. 1 µL of each sample or D1000 Ladder was mixed with 3 µL Agilent D1000 Tapestation Buffer. Gels are imaged at roughly 60% saturation. A) All libraries were amplified for 15 cycles. B) Libraries subjected to reconditioning PCR display anticipated size distribution (~350 bp). Reconditioning was performed by taking 16 uL of overamplified library and subjecting it to round of PCR amplification, followed by a 0.8x bead based cleanup.



S16: 5609 B1 pSer7 Libraries Visualized on an Agilent 2200 Tapestation. 1 µL of each sample or D1000 Ladder was mixed with 3 µL Agilent D1000 Tapestation Buffer. Gels are imaged at roughly 60% saturation. A) All libraries were amplified for 15 cycles. B) Libraries subjected to reconditioning PCR display anticipated size distribution (~350 bp). Reconditioning was performed by taking 16 uL of overamplified library and subjecting it to round of PCR amplification, followed by a 0.8x bead based cleanup.



S17: 5609 B1 pSer2 Libraries Visualized on an Agilent 2200 Tapestation. 1 µL of each sample or D1000 Ladder was mixed with 3 µL Agilent D1000 Tapestation Buffer. Gels are imaged at roughly 60% saturation. A) All libraries were amplified for 15 cycles. B) Libraries subjected to reconditioning PCR display anticipated size distribution (~350 bp). Reconditioning was performed by taking 16 uL of overamplified library and subjecting it to round of PCR amplification, followed by a 0.8x bead based cleanup.



S18:-5609 B2 Libraries Visualized on an Agilent 2200 Tapestation. 1 µL of each sample or D1000 Ladder was mixed with 3 µL Agilent D1000 Tapestation Buffer. Gels are imaged at roughly 60% saturation. A) All libraries were amplified for 14 cycles. B) Libraries subjected to reconditioning PCR display anticipated size distribution (~350 bp). Reconditioning was performed by taking 19 uL of overamplified library and subjecting it to round of PCR amplification, followed by a 0.8x bead based cleanup. S7 1-4 and S5-3were not reconditioned, but were subjected to an additional 0.8x bead based cleanup step to removeadapter dimers. N1I-4I samples shown here were imaged following the standard library preparation and subjected to reconditioning as described above (not shown). 1= DMSO treatment, 2 = SY5609, 3=KB0742, 4=SY/KB

Set to pool								
Tube ID	**Replicate**	**Treatment**	**IP**	**Index**	**Qubit (ng/uL)**	**Dilution (ng/uL)**	**Volume used (uL)**	Total ng
A	1	DMSO	NTD	UDI26	39.40	3.94	5.08	20
B	2	DMSO	pSer5	UDI19	35.60	3.56	3.37	12
C	1	SY-351	NTD	UDI34	50.00	5.00	4.00	20
D	2	SY-351	pSer5	UDI27	32.40	3.24	3.70	12
E	1	SY-351/KB-0742	NTD	UDI42	16.10	16.10	1.24	20
F	2	SY-351/KB-0742	pSer5	UDI35	9.96	9.96	1.20	12
G	1	DMSO	Input (from NTD)	UDI50	35.80	3.58	1.12	4
H	1	SY-351	Input (from NTD)	UDI58	34.60	3.46	1.16	4
I	1	SY-351/KB-0742	Input (from NTD)	UDI66	36.20	3.62	1.10	4
						Pool volume	**21.98**	
						Pool concentration (ng/ul):	3.7	
						Average Size (bp)	352	
						nM Concentration	15.92630854	
						Desired Conc (nM)	10	
						Desired Vol (uL)	30	
						Vol pool used (uL)	19.46	
						Vol buffer used (uL)	11.54	
						Final concentration (ng/uL)	2.18	
						Final concentration (nM)	9.383608815	

S19: SY-351 T1 Library Pooling Scheme. Library concentrations were determined via Qubit using the Qubit dsDNA High Sensitivity Kit and Standards. Target dilutions were calculated via desired total volume, sample ratio, and total pool concentration. The final pool was visualized on an Agilent 2200 Tapestaion to determine the average size, while Qubit was used to determine total concentration. This information was used to calculate a nM concentration for the pool, which was then used to create the final pool that contained 30 uL at a target 10 nM concentration.

Set to pool									
Tube ID	**Replicate**	**Treatment**	**IP**	**Index**	**Qubit (ng/uL)**	**Dilution (ng/uL)**	**Volume used (uL)**	Total ng	Peak
A	2	DMSO	NTD	UDI01	33.20	3.32	6.02	20	325
B	2	SY-351	NTD	UDI17	42.80	4.28	4.67	20	373
C	2	SY-351/KB-0742	NTD	UDI25	12.90	12.90	1.55	20	338
D	1	DMSO	pSer5	UDI04	25.40	2.54	4.72	12	344
E	1	SY-351	pSer5	UDI20	6.72	6.72	1.79	12	348
F	1	SY-351/KB-0742	pSer5	UDI28	22.00	2.20	5.45	12	366
G	1	DMSO	Input (from pSer5)	UDI36	26.20	2.62	1.53	4	357
H	1	SY-351	Input (from pSer5)	UDI44	30.20	3.02	1.32	4	356
I	1	SY-351/KB-0742	Input (from pSer5)	UDI52	16.60	1.66	2.41	4	361
						Pool volume	**29.47**		
						Pool concentration (ng/ul):	3.664381996		
						Average Size (bp)	345		
						nM Concentration	16.09302589		
						Desired nM Conc	10		
						Desired Vol (uL)	30		
						Vol Pool Needed	18.64		
						Vol Buffer Needed	11.36		

S20: SY351 T2 Pooling Scheme. Library concentrations were determined via Qubit using the Qubit dsDNA High Sensitivity Kit and Standards. Target dilutions were calculated via desired total volume, sample ratio, and total pool concentration. The size of the final pool was determined by taking a weighted average of individual library sizes, while Qubit was used to determine total concentration. This information was used to calculate a nM concentration for the pool, which was then used to create the final pool that contained 30 uL at a target 10 nM concentration. Subsequent analysis of the pool on the Tapestation revealed the average size to be 349 bp, resulting in a < 1 µL error.

Set to pool											
Tube ID	**Replicate**	**Treatment**	**IP**	**Index**	**Stock Qbit (ng/uL)**	**Dilution (ng/uL)**	**Volume used (uL)**	Total ng	Peak Size		
A	1	DMSO	NTD	UDI65	23.00	2.30	4.35	10	370		
B	1	SY-5609	NTD	UDI66	55.00	5.50	1.82	10	363		
C	1	KB-0742	NTD	UDI67	25.80	2.58	3.88	10	360		
D	1	SY/KB	NTD	UDI68	39.60	3.96	2.53	10	350		
E	1	DMSO	pSer5	UDI81	13.00	1.30	4.62	6	375		
F	1	SY-5609	pSer5	UDI82	11.40	5.70	1.05	6	362		
G	1	KB-0742	pSer5	UDI83	22.80	2.28	2.63	6	380		
H	1	SY/KB	pSer5	UDI84	33.20	3.32	1.81	6	377		
I	1	DMSO	pSer2	UDI89	12.70	3.18	1.89	6	342		
J	1	SY-5609	pSer2	UDI90	37.80	3.78	1.59	6	341		
K	1	KB-0742	pSer2	UDI91	25.60	2.56	2.34	6	367		
L	1	SY/KB	pSer2	UDI92	41.40	4.14	1.45	6	369		
M	1	DMSO	pSer7	UDI73	26.20	2.62	2.29	6	374	**Pool volume**	**41.81**
N	1	SY-5609	pSer7	UDI74	25.40	2.54	2.36	6	371	Pool concentration (ng/ul):	2.870389709
O	1	KB-0742	pSer7	UDI75	22.60	5.65	1.06	6	372	Average Size (bp)	344
P	1	SY/KB	pSer7	UDI76	19.70	4.93	1.22	6	365	nM Concentration	12.6426608
Q	1	DMSO	Input (from pSer7)	UDI77	33.40	1.67	1.20	2	360	Desired nM Concentration	10
R	1	SY-5609	Input (from pSer7)	UDI78	34.60	1.73	1.16	2	359	Desired Pool Volume (uL)	30
S	1	KB-0742	Input (from pSer7)	UDI79	44.20	1.47	1.36	2	358	Vol Pool Needed (uL)	23.73
T	1	SY/KB	Input (from pSer7)	UDI80	32.80	1.64	1.22	2	349	Vol Buffer Needed (uL)	6.27

S21: SY5609 B1 Pooling Scheme. Library concentrations were determined via Qubit using the Qubit dsDNA High Sensitivity Kit and Standards. Target dilutions were calculated via desired total volume, sample ratio, and total pool concentration. The final pool was visualized on an Agilent 2200 Tapestaion to determine the average size, while Qubit was used to determine total concentration. This information was used to calculate a nM concentration for the pool, which was then used to create the final pool that contained 30 uL at a target 10 nM concentration.

Set to pool										
Tube ID	**Replicate**	**Treatment**	**IP**	**Index**	**Stock Qbit (ng/uL)**	**Dilution (ng/uL)**	**Volume used (uL)**	Total ng		
A	2	DMSO	NTD	UDI65	8.32	8.32	1.20	10		
B	2	SY-5609	NTD	UDI66	28.20	2.82	3.55	10		
C	2	KB-0742	NTD	UDI67	28.80	2.88	3.47	10		
D	2	SY/KB	NTD	UDI68	51.40	5.14	1.95	10		
E	2	DMSO	pSer5	UDI81	4.04	4.04	1.49	6		
F	2	SY-5609	pSer5	UDI82	42.60	4.26	1.41	6		
G	2	KB-0742	pSer5	UDI83	8.82	4.41	1.36	6		
H	2	SY/KB	pSer5	UDI84	71.40	1.43	4.20	6		
I	2	DMSO	pSer2	UDI89	6.28	3.14	1.91	6		
J	2	SY-5609	pSer2	UDI90	19.60	1.96	3.06	6		
K	2	KB-0742	pSer2	UDI91	34.80	3.48	1.72	6	**Pool volume**	**40.32**
L	2	SY/KB	pSer2	UDI92	31.40	3.14	1.91	6	Pool concentration (ng/ul):	3.3
M	2	DMSO	pSer7	UDI73	4.86	4.86	1.23	6	Average Size (bp)	365
N	2	SY-5609	pSer7	UDI74	2.82	2.82	2.13	6	nM Concentration	13.69863014
O	2	KB-0742	pSer7	UDI75	4.80	4.80	1.25	6	Desired nM Concentration	10
P	2	SY/KB	pSer7	UDI76	3.18	3.18	1.89	6	Desired Pool Volume (uL)	30
Q	2	DMSO	Input (from NTD)	UDI69	57.80	1.16	1.73	2	Vol Pool Needed (uL)	21.90
R	2	SY-5609	Input (from NTD)	UDI70	65.60	1.31	1.52	2	Vol Buffer Needed (uL)	8.10
S	2	KB-0742	Input (from NTD)	UDI71	59.80	1.20	1.67	2		
T	2	SY/KB	Input (from NTD)	UDI72	48.00	1.20	1.67	2		

S22: SY5609 B2 Pooling Scheme. Library concentrations were determined via Qubit using the Qubit dsDNA High Sensitivity Kit and Standards. Target dilutions were calculated via desired total volume, sample ratio, and total pool concentration. The final pool was visualized on an Agilent 2200 Tapestaion to determine the average size, while Qubit was used to determine total concentration. This information was used to calculate a nM concentration for the pool, which was then used to create the final pool that contained 30 uL at a target 10 nM concentration.

Lane	Sample	Barcode sequence	PF Clusters	% of the	% Perfect	% One mismatch	Yield (Mbases)	% PF	% >= Q30	Mean Quality
				lane	barcode	barcode		Clusters	bases	Score
3	DMSO_I	TCTGGCGA+GTTGAATT	6,521,930	0.24	92.52	7.48	1,970	100	81.91	33.3
3	DMSO_N1	GTATAACA+GTGAAGCA	35,236,073	1.31	92.56	7.44	10,641	100	85.89	34.25
3	DMSO_S2	ATTCCTCT+AATATCGC	17,328,967	0.64	85.98	14.02	5,233	100	84.44	33.94
3	SY351_I	TCCTTGGT+GTCGATTC	3,151,394	0.12	92.12	7.88	952	100	82.02	33.42
3	SY351_KB0742_I	AGGTTATC+GGTATGAT	5,818,052	0.22	93.04	6.96	1,757	100	81.75	33.3
3	SY351_KB0742_N1	CCAGTTAG+ATGACAGG	32,661,334	1.21	87.74	12.26	9,864	100	82.21	33.4
3	SY351_KB0742_S2	CATGCTTA+TAATTCAT	18,772,604	0.7	91.42	8.58	5,669	100	86.42	34.38
3	SY351_N1	TTCGCTGA+AATATGCG	34,373,907	1.28	88	12	10,381	100	83.15	33.61
3	SY351_S2	TGTCGGAT+AAGCCTGA	14,878,720	0.55	91.89	8.11	4,493	100	87.49	34.63

S23: QC Scores from Demultiplexed SY-351 T1 Data-Run 1. All data shown was produced by the Gates Institute Genomics Core.

Lane	Sample	Barcode sequence	PF Clusters	% of the	% Perfect	% One mismatch	Yield (Mbases)	% PF	% >= Q30	Mean Quality
				lane	barcode	barcode		Clusters	bases	Score
3	DMSO_I	TCTGGCGA+GTTGAATT	4,550,073	0.19	89.99	10.01	1,374	100	77.12	32.03
3	DMSO_N1	GTATAACA+GTGAAGCA	24,709,791	1.01	92.45	7.55	7,462	100	81.1	32.99
3	DMSO_S2	ATTCCTCT+AATATCGC	11,818,558	0.48	88.79	11.21	3,569	100	79.61	32.66
3	SY351_I	TCCTTGGT+GTCGATTC	2,136,367	0.09	91.57	8.43	645	100	77.84	32.28
3	SY351_KB0742_I	AGGTTATC+GGTATGAT	4,206,391	0.17	92.43	7.57	1,270	100	77.11	32.07
3	SY351_KB0742_N1	CCAGTTAG+ATGACAGG	23,293,233	0.95	88.72	11.28	7,035	100	77.52	32.15
3	SY351_KB0742_S2	CATGCTTA+TAATTCAT	13,464,917	0.55	92.33	7.67	4,066	100	81.55	33.1
3	SY351_N1	TTCGCTGA+AATATGCG	24,070,712	0.98	92.66	7.34	7,269	100	78.22	32.31
3	SY351_S2	TGTCGGAT+AAGCCTGA	10,589,397	0.43	94.69	5.31	3,198	100	82.74	33.38

S24: QC Scores from Demultiplexed SY-351 T1 Data-Run 2. All data shown was produced by the Gates Institute Genomics Core.

Lane	Sample	Barcode sequence	PF Clusters	% of the	% Perfect	% One mismatch	Yield (Mbases)	% PF	% >= Q30	Mean Quality
				lane	barcode	barcode		Clusters	bases	Score
4	351_I1	AAGTAGAG+CAGTCGAT	7,131,030	0.27	82.65	17.35	2,154	100	84.99	34.1
4	351_I2	GGTCCAGA+AAGAACCG	5,507,496	0.21	93.94	6.06	1,663	100	87.33	34.66
4	351_I4	ATTATGTT+GAATACAG	5,309,949	0.2	92.23	7.77	1,604	100	88.2	34.85
4	351_N1	GTAACATC+AATCGCTG	45,444,778	1.73	95.03	4.97	13,724	100	83.81	33.81
4	351_N2	TGCTGCTG+ATGAGGAC	33,701,566	1.29	93.42	6.58	10,178	100	86.14	34.37
4	351_N4	ACACGATC+ATGGTATT	62,147,267	2.37	92.92	7.08	18,768	100	77.86	32.49
4	351_S1	AATGTTCT+AGGTGACT	35,589,053	1.36	94.26	5.74	10,748	100	86.29	34.4
4	351_S2	CAACTCTC+GAGAAGAT	35,908,460	1.37	91.62	8.38	10,844	100	79.98	32.98
4	351_S4	AGGATCTA+CGAAGTTC	25,153,288	0.96	93.01	6.99	7,596	100	88.78	34.96

S25: QC Scores from Demultiplexed SY-351 T2 Data. All data shown was produced by the Gates Institute Genomics Core.

Lane	Sample	Barcode sequence	PF Clusters	% of the	% Perfect	% One mismatch	Yield (Mbases)	% PF	% >= Q30	Mean Quality
				lane	barcode	barcode		Clusters	bases	Score
4	5609_B1_21	ACAGTTGA+ATCGGAGC	13,395,685	0.67	71.8	28.2	4,045	100	84.07	33.96
4	5609_B1_22	TTGTCTAT+GATTCAAG	18,566,939	0.93	53.99	46.01	5,607	100	79.14	32.81
4	5609_B1_23	CGCTATGT+TATGCGGA	7,961,429	0.4	78.57	21.43	2,404	100	76.01	32.09
4	5609_B1_24	TTAATCAG+CAGATTGG	3,918,603	0.2	81.5	18.5	1,183	100	64.13	29.3
4	5609_B1_51	GACCGTTG+TGTGATGG	7,221,931	0.36	68.07	31.93	2,181	100	76.46	32.19
4	5609_B1_52	AACAATGG+GCTAACTA	6,080,609	0.3	82.44	17.56	1,836	100	74.77	31.77
4	5609_B1_53	AGGTGCGA+GCCAGAAG	2,994,276	0.15	77.72	22.28	904	100	72.07	31.14
4	5609_B1_54	AGGTCGCA+AATTGTGC	1,784,200	0.09	61.71	38.29	539	100	59.92	28.27
4	5609_B1_71	CTACCAGG+ATCAGTTG	6,368,783	0.32	74.72	25.28	1,923	100	65.15	29.45
4	5609_B1_72	ACTGTATC+TAATAGCA	10,571,705	0.53	65.92	34.08	3,193	100	71.17	30.91
4	5609_B1_73	CTGTGGCG+GGCTAGTG	3,053,223	0.15	81.55	18.45	922	100	60.95	28.56
4	5609_B1_74	TGTAATCA+TGGAGATT	5,339,662	0.27	64.06	35.94	1,613	100	62.12	28.8
4	5609_B1_I1	TTATATCT+GTGCAGAC	2,568,156	0.13	52.13	47.87	776	100	67.87	30.21
4	5609_B1_I2	GCCGCAAC+AGACATGA	2,473,730	0.12	81.31	18.69	747	100	66.48	29.84
4	5609_B1_I3	TGTAACTC+AACTGTCG	2,476,548	0.12	74.42	25.58	748	100	69.65	30.59
4	5609_B1_I4	CTGCGGAT+AGATAACC	2,796,200	0.14	81.14	18.86	844	100	69.16	30.47
4	5609_B1_N1	TTACGCAC+AGATGGTC	9,423,432	0.47	73.64	26.36	2,846	100	73.21	31.4
4	5609_B1_N2	AGGTTATC+GGTATGAT	11,099,890	0.55	71.47	28.53	3,352	100	70.76	30.84
4	5609_B1_N3	TCGCCTTG+GAATCAGC	10,504,516	0.52	84.15	15.85	3,172	100	65.85	29.7
4	5609_B1_N4	CCAGAGCT+GAAGTTCG	9,612,429	0.48	80.98	19.02	2,903	100	63.86	29.25

S26: QC Scores from Demultiplexed SY-5609 B1 Data. All data shown was produced by the Gates Institute Genomics Core.

Lane	Sample	Barcode sequence	PF Clusters	% of the	% Perfect	% One mismatch	Yield (Mbases)	% PF	% >= Q30	Mean Quality
				lane	barcode	barcode		Clusters	bases	Score
2	5609_B2_21	ACAGTTGA+ATCGGAGC	5,019,110	0.24	78.28	21.72	1,516	100	81.9	33.45
2	5609_B2_22	TTGTCTAT+GATTCAAG	16,762,159	0.81	65.56	34.44	5,062	100	82.38	33.57
2	5609_B2_23	CGCTATGT+TATGCGGA	5,199,907	0.25	85.6	14.4	1,570	100	80.73	33.19
2	5609_B2_24	TTAATCAG+CAGATTGG	13,359,224	0.65	89.64	10.36	4,034	100	73.38	31.49
2	5609_B2_51	GACCGTTG+TGTGATGG	13,131,714	0.64	81.42	18.58	3,966	100	85.54	34.31
2	5609_B2_52	AACAATGG+GCTAACTA	1,541,320	0.07	81.39	18.61	465	100	71.54	31.04
2	5609_B2_53	AGGTGCGA+GCCAGAAG	14,397,531	0.7	86.76	13.24	4,348	100	79.27	32.86
2	5609_B2_54	AGGTCGCA+AATTGTGC	3,276,502	0.16	64.62	35.38	990	100	66.6	29.89
2	5609_B2_71	CTACCAGG+ATCAGTTG	28,703,228	1.39	77.86	22.14	8,668	100	83.71	33.89
2	5609_B2_72	ACTGTATC+TAATAGCA	38,369,111	1.86	78.53	21.47	11,587	100	84.01	33.95
2	5609_B2_73	CTGTGGCG+GGCTAGTG	43,222,176	2.1	87.89	12.11	13,053	100	86.75	34.6
2	5609_B2_74	TGTAATCA+TGGAGATT	39,576,632	1.92	82.39	17.61	11,952	100	85.92	34.4
2	5609_B2_I1	TTATATCT+GTGCAGAC	42	0	45.24	54.76	0	100	67.21	30.03
2	5609_B2_I2	GCCGCAAC+AGACATGA	59	0	83.05	16.95	0	100	64.58	29.4
2	5609_B2_I3	TGTAACTC+AACTGTCG	17	0	52.94	47.06	0	100	61.2	28.46
2	5609_B2_I4	CTGCGGAT+AGATAACC	105	0	79.05	20.95	0	100	64.96	29.48
2	5609_B2_N1	TTACGCAC+AGATGGTC	19,388,581	0.94	75.77	24.23	5,855	100	79.75	32.94
2	5609_B2_N2	AGGTTATC+GGTATGAT	10,583,386	0.51	74.09	25.91	3,196	100	75.58	32
2	5609_B2_N3	TCGCCTTG+GAATCAGC	43,348,700	2.1	90.8	9.2	13,091	100	82.86	33.68
2	5609_B2_N4	CCAGAGCT+GAAGTTCG	8,151,884	0.4	84.18	15.82	2,462	100	71.75	31.13

S27: QC Scores from Demultiplexed SY-5609 B2 Run 1 Data. All data shown was produced by the Gates Institute Genomics Core.

Lane	Sample	Barcode sequence	PF Clusters	% of the	% Perfect	% One mismatch	Yield (Mbases)	% PF	% >= Q30	Mean Quality
				lane	barcode	barcode		Clusters	bases	Score
1	5609_B2_21	ACAGTTGA+ATCGGAGC	2,781,636	0.12	82.33	17.67	840	100	82.09	33.47
1	5609_B2_22	TTGTCTAT+GATTCAAG	9,630,398	0.41	85.96	14.04	2,908	100	81.32	33.27
1	5609_B2_23	CGCTATGT+TATGCGGA	2,933,975	0.12	87.87	12.13	886	100	79.95	32.97
1	5609_B2_24	TTAATCAG+CAGATTGG	7,270,099	0.31	86.48	13.52	2,196	100	72.29	31.09
1	5609_B2_51	GACCGTTG+TGTGATGG	7,634,708	0.32	83.54	16.46	2,306	100	84.66	34.09
1	5609_B2_52	AACAATGG+GCTAACTA	783,872	0.03	85.73	14.27	237	100	71.29	30.85
1	5609_B2_53	AGGTGCGA+GCCAGAAG	7,945,347	0.34	87.54	12.46	2,399	100	78.44	32.57
1	5609_B2_54	AGGTCGCA+AATTGTGC	1,673,133	0.07	69.01	30.99	505	100	63.51	29.09
1	5609_B2_71	CTACCAGG+ATCAGTTG	16,806,259	0.71	76.44	23.56	5,075	100	83.13	33.7
1	5609_B2_72	ACTGTATC+TAATAGCA	22,774,309	0.97	83.41	16.59	6,878	100	83.14	33.69
1	5609_B2_73	CTGTGGCG+GGCTAGTG	24,718,534	1.05	86.62	13.38	7,465	100	85.81	34.36
1	5609_B2_74	TGTAATCA+TGGAGATT	22,971,348	0.98	90.04	9.96	6,937	100	85.58	34.28
1	5609_B2_I1	TTATATCT+GTGCAGAC	32	0	78.13	21.88	0	100	70	30.6
1	5609_B2_I2	GCCGCAAC+AGACATGA	36	0	94.44	5.56	0	100	61.42	28.51
1	5609_B2_I3	TGTAACTC+AACTGTCG	5	0	100	NaN	0	100	58.41	27.91
1	5609_B2_I4	CTGCGGAT+AGATAACC	60	0	63.33	36.67	0	100	63.89	29.33
1	5609_B2_N1	TTACGCAC+AGATGGTC	10,974,633	0.47	85.01	14.99	3,314	100	79.34	32.79
1	5609_B2_N2	AGGTTATC+GGTATGAT	5,638,712	0.24	80.66	19.34	1,703	100	75.33	31.86
1	5609_B2_N3	TCGCCTTG+GAATCAGC	25,005,132	1.06	93.4	6.6	7,552	100	81.85	33.41
1	5609_B2_N4	CCAGAGCT+GAAGTTCG	4,220,536	0.18	84.12	15.88	1,275	100	70.27	30.7

S28: QC Scores from Demultiplexed SY-5609 B2 Run 2 Data. All data shown was produced by the Gates Institute Genomics Core.

S29: Table Displaying List of Top 263 Highly Expressed Genes in HCT116 Cells[163]

Symbol	Name	Standardized Value
DNAJC27-AS1	DNAJC27 antisense RNA 1	3.40796
OR51B2	olfactory receptor, family 51, subfamily B, member 2 (gene/pseudogene)	3.40793
ABHD10	abhydrolase domain containing 10	3.26764
SSFA2	sperm specific antigen 2	3.05687
DNAAF3	dynein, axonemal, assembly factor 3	3.01933
OR51I1	olfactory receptor, family 51, subfamily I, member 1	2.88548
RCC1	regulator of chromosome condensation 1	2.86949
RAC3	ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3)	2.86485
VTI1A	vesicle transport through interaction with t-SNAREs 1A	2.85041
MCMBP	minichromosome maintenance complex binding protein	2.78464
PDP1	pyruvate dehyrogenase phosphatase catalytic subunit 1	2.6297
CTU2	cytosolic thiouridylase subunit 2 homolog (S. pombe)	2.60844
THRAP3	thyroid hormone receptor associated protein 3	2.50764
GRIN2B	glutamate receptor, ionotropic, N-methyl D-aspartate 2B	2.49248
CPOX	coproporphyrinogen oxidase	2.40376
EIF5A	eukaryotic translation initiation factor 5A	2.35663
NPM3	nucleophosmin/nucleoplasmin 3	2.32549
NOLC1	nucleolar and coiled-body phosphoprotein 1	2.31865
AP1S3	adaptor-related protein complex 1, sigma 3 subunit	2.30834
EXOSC9	exosome component 9	2.30832
CHUK	conserved helix-loop-helix ubiquitous kinase	2.29554
TMEM180	transmembrane protein 180	2.28973
TCF25	transcription factor 25 (basic helix-loop-helix)	2.28537
MAGEB2	melanoma antigen family B2	2.23111
HYALP1	hyaluronoglucosaminidase pseudogene 1	2.21344
ZDHHC6	zinc finger, DHHC-type containing 6	2.20582
EZR	ezrin	2.1996
OR51B4	olfactory receptor, family 51, subfamily B, member 4	2.19879
MTAP	methylthioadenosine phosphorylase	2.19172
JMJD8	jumonji domain containing 8	2.18472
WNT16	wingless-type MMTV integration site family, member 16	2.13763
HSP90AB1	heat shock protein 90kDa alpha (cytosolic), class B member 1	2.11774

SFR1	SWI5-dependent recombination repair 1	2.10432
ZIC5	Zic family member 5	2.08024
EMC8	ER membrane protein complex subunit 8	2.07791
ULBP3	UL16 binding protein 3	2.07196
HPS6	Hermansky-Pudlak syndrome 6	2.05585
XAB2	XPA binding protein 2	2.04084
C16ORF46	chromosome 16 open reading frame 46	2.03581
KCNQ2	potassium channel, voltage gated KQT-like subfamily Q, member 2	2.03065
KITLG	KIT ligand	2.00221
PSMD12	proteasome (prosome, macropain) 26S subunit, non-ATPase, 12	1.99929
MRPL12	mitochondrial ribosomal protein L12	1.99907
GABRD	gamma-aminobutyric acid (GABA) A receptor, delta	1.99464
MRPS23	mitochondrial ribosomal protein S23	1.99396
M1AP	meiosis 1 associated protein	1.97687
PHF5A	PHD finger protein 5A	1.96931
PGAM5	phosphoglycerate mutase family member 5	1.96174
KLHDC4	kelch domain containing 4	1.9545
MRPL2	mitochondrial ribosomal protein L2	1.94673
NTSR1	neurotensin receptor 1 (high affinity)	1.93446
TRIB1	tribbles pseudokinase 1	1.90636
SYTL3	synaptotagmin-like 3	1.89962
MCAT	malonyl CoA:ACP acyltransferase (mitochondrial)	1.89756
ATG12	autophagy related 12	1.89738
KRR1	KRR1, small subunit (SSU) processome component, homolog (yeast)	1.89617
FAM175B	family with sequence similarity 175, member B	1.89503
TRUB1	TruB pseudouridine (psi) synthase family member 1	1.89405
OR5K1	olfactory receptor, family 5, subfamily K, member 1	1.89273
LYZL6	lysozyme-like 6	1.89086
ST7-OT4	ST7 overlapping transcript 4	1.88799
OR51B5	olfactory receptor, family 51, subfamily B, member 5	1.88441
TIAL1	TIA1 cytotoxic granule-associated RNA binding protein-like 1	1.8837
NME1	NME/NM23 nucleoside diphosphate kinase 1	1.87764
WDR43	WD repeat domain 43	1.86736
MRPL38	mitochondrial ribosomal protein L38	1.86709
MRPL1	mitochondrial ribosomal protein L1	1.86586

CCDC124	coiled-coil domain containing 124	1.86576
C1QBP	complement component 1, q subcomponent binding protein	1.86036
PPP1R2P9	protein phosphatase 1, regulatory (inhibitor) subunit 2 pseudogene 9	1.85935
PHB	prohibitin	1.85572
TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)	1.85282
RPF1	ribosome production factor 1 homolog (S. cerevisiae)	1.85018
PDCD11	programmed cell death 11	1.84035
GEMIN4	gem (nuclear organelle) associated protein 4	1.83662
STK39	serine threonine kinase 39	1.83062
FRMD5	FERM domain containing 5	1.83
RFC5	replication factor C (activator 1) 5, 36.5kDa	1.82565
XRCC4	X-ray repair complementing defective repair in Chinese hamster cells 4	1.82133
NPW	neuropeptide W	1.81811
LLPH	LLP homolog, long-term synaptic facilitation (Aplysia)	1.8147
PTH	parathyroid hormone	1.80361
RRP9	ribosomal RNA processing 9, small subunit (SSU) processome component, homolog (yeast)	1.79385
POP1	processing of precursor 1, ribonuclease P/MRP subunit (S. cerevisiae)	1.7932
NAF1	nuclear assembly factor 1 ribonucleoprotein	1.78872
MPP7	membrane protein, palmitoylated 7 (MAGUK p55 subfamily member 7)	1.78855
RNF126	ring finger protein 126	1.78802
DOHH	deoxyhypusine hydroxylase/monooxygenase	1.78732
TRPM2	transient receptor potential cation channel, subfamily M, member 2	1.78387
ZNF205	zinc finger protein 205	1.77083
TOMM22	translocase of outer mitochondrial membrane 22 homolog (yeast)	1.76779
CCDC59	coiled-coil domain containing 59	1.74641
C10ORF2	chromosome 10 open reading frame 2	1.74492
SH3RF1	SH3 domain containing ring finger 1	1.74412
SPATA2L	spermatogenesis associated 2-like	1.73448
EDDM3B	epididymal protein 3B	1.73444
WBP4	WW domain binding protein 4	1.73323
MAPK8	mitogen-activated protein kinase 8	1.73269
EIF4A3	eukaryotic translation initiation factor 4A3	1.72924
PRKACB	protein kinase, cAMP-dependent, catalytic, beta	1.71315
RNMTL1	RNA methyltransferase like 1	1.70877
ELAC2	elaC ribonuclease Z 2	1.70853

HBE1	hemoglobin, epsilon 1	1.70612
GCAT	glycine C-acetyltransferase	1.7053
ICT1	immature colon carcinoma transcript 1	1.6997
TMEM52	transmembrane protein 52	1.69862
BMS1	BMS1 ribosome biogenesis factor	1.69794
BCCIP	BRCA2 and CDKN1A interacting protein	1.69586
SHISA8	shisa family member 8	1.69129
ASIP	agouti signaling protein	1.68265
JMJD6	jumonji domain containing 6	1.67904
AKAP12	A kinase (PRKA) anchor protein 12	1.67582
HNRNPL	heterogeneous nuclear ribonucleoprotein L	1.66964
KRTAP3-3	keratin associated protein 3-3	1.66616
PPRC1	peroxisome proliferator-activated receptor gamma, coactivator-related 1	1.663
SLC34A1	solute carrier family 34 (type II sodium/phosphate cotransporter), member 1	1.66027
C10ORF76	chromosome 10 open reading frame 76	1.6535
METAP1	methionyl aminopeptidase 1	1.65204
PRPF4	pre-mRNA processing factor 4	1.6489
RSPH3	radial spoke 3 homolog (Chlamydomonas)	1.6429
POLR1C	polymerase (RNA) I polypeptide C, 30kDa	1.64124
KRT222	keratin 222, type II	1.64116
MRPL44	mitochondrial ribosomal protein L44	1.6397
TBL3	transducin (beta)-like 3	1.63846
NOC4L	nucleolar complex associated 4 homolog (S. cerevisiae)	1.63398
LTB4R2	leukotriene B4 receptor 2	1.63154
NLRP6	NLR family, pyrin domain containing 6	1.63151
PRR22	proline rich 22	1.63075
RAD51C	RAD51 paralog C	1.62792
STRA13	stimulated by retinoic acid 13	1.62781
SAMD3	sterile alpha motif domain containing 3	1.62744
CTR9	CTR9, Paf1/RNA polymerase II complex component	1.62417
CTRL	chymotrypsin-like	1.61662
C1ORF174	chromosome 1 open reading frame 174	1.61328
CYB5B	cytochrome b5 type B (outer mitochondrial membrane)	1.60846
SLC19A1	solute carrier family 19 (folate transporter), member 1	1.60377
TSEN54	TSEN54 tRNA splicing endonuclease subunit	1.60153

POLR3G	polymerase (RNA) III (DNA directed) polypeptide G (32kD)	1.60113
PWP2	PWP2 periodic tryptophan protein homolog (yeast)	1.60074
POLR1E	polymerase (RNA) I polypeptide E, 53kDa	1.60067
KLHL17	kelch-like family member 17	1.59774
GEMIN5	gem (nuclear organelle) associated protein 5	1.59458
CATSPER3	cation channel, sperm associated 3	1.59074
SZRD1	SUZ RNA binding domain containing 1	1.58927
KRTAP2-2	keratin associated protein 2-2	1.58413
MSX1	msh homeobox 1	1.58391
CLDND1	claudin domain containing 1	1.57811
GRPEL1	GrpE-like 1, mitochondrial (E. coli)	1.57156
LIFR-AS1	LIFR antisense RNA 1	1.5712
TSR1	TSR1, 20S rRNA accumulation, homolog (S. cerevisiae)	1.57109
ERVMER34-1	endogenous retrovirus group MER34, member 1	1.56801
MTERF3	mitochondrial transcription termination factor 3	1.56572
MRPL10	mitochondrial ribosomal protein L10	1.56082
KXD1	KxDL motif containing 1	1.55951
NOL11	nucleolar protein 11	1.55263
ARMC7	armadillo repeat containing 7	1.54988
SNORA61	small nucleolar RNA, H/ACA box 61	1.54949
SNHG9	small nucleolar RNA host gene 9	1.54889
HAS3	hyaluronan synthase 3	1.54675
CACUL1	CDK2-associated, cullin domain 1	1.53913
RNF141	ring finger protein 141	1.5387
MRTO4	mRNA turnover 4 homolog (S. cerevisiae)	1.53546
MRRF	mitochondrial ribosome recycling factor	1.53546
METTL21C	methyltransferase like 21C	1.52903
LINC00477	long intergenic non-protein coding RNA 477	1.52776
MRPL43	mitochondrial ribosomal protein L43	1.52726
ADAP2	ArfGAP with dual PH domains 2	1.52625
SEC23IP	SEC23 interacting protein	1.52613
STIP1	stress-induced phosphoprotein 1	1.52537
MAP2K4	mitogen-activated protein kinase kinase 4	1.51973
ACTR5	ARP5 actin-related protein 5 homolog (yeast)	1.5178
CHMP1A	charged multivesicular body protein 1A	1.51645

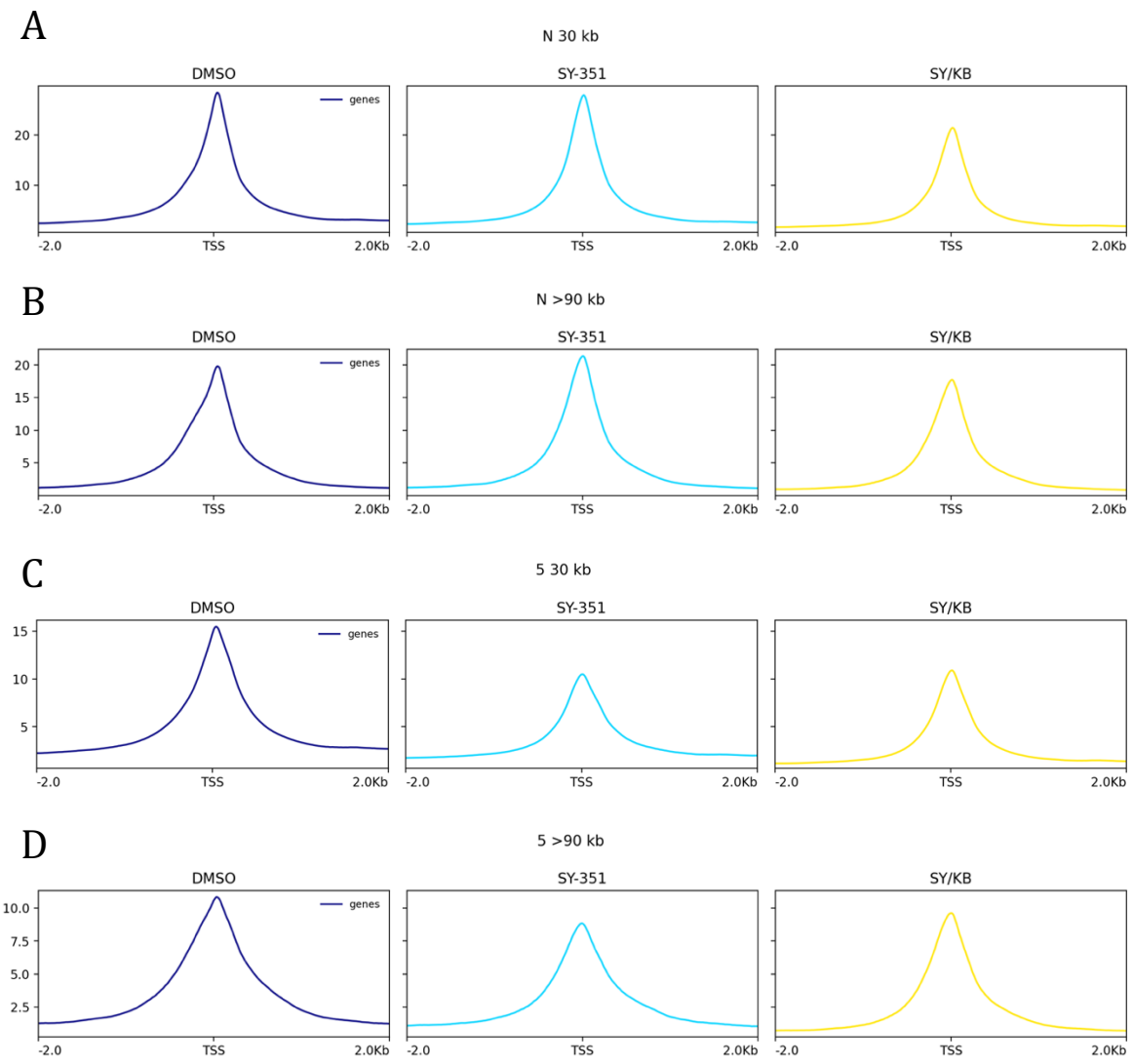
RPUSD1	RNA pseudouridylate synthase domain containing 1	1.51645
RFPL2	ret finger protein-like 2	1.51377
PPID	peptidylprolyl isomerase D	1.51227
NEIL2	nei endonuclease VIII-like 2 (E. coli)	1.51081
MRPL42	mitochondrial ribosomal protein L42	1.5107
HS3ST1	heparan sulfate (glucosamine) 3-O-sulfotransferase 1	1.50838
DYNLL2	dynein, light chain, LC8-type 2	1.5068
HIST1H1E	histone cluster 1, H1e	1.50618
UTP11L	UTP11-like, U3 small nucleolar ribonucleoprotein (yeast)	1.50501
UTP15	UTP15, U3 small nucleolar ribonucleoprotein, homolog (S. cerevisiae)	1.5046
NUDC	nudC nuclear distribution protein	1.5033
MPHOSPH6	M-phase phosphoprotein 6	1.50282
NOP56	NOP56 ribonucleoprotein	1.50133
LYAR	Ly1 antibody reactive	1.50122
GAL	galanin/GMAP prepropeptide	1.49715
LTV1	LTV1 ribosome biogenesis factor	1.4964
TOP1	topoisomerase (DNA) I	1.49583
YRDC	yrdC N(6)-threonylcarbamoyltransferase domain containing	1.49247
GTSF1	gametocyte specific factor 1	1.48511
SEC14L4	SEC14-like 4 (S. cerevisiae)	1.48496
DPH7	diphthamide biosynthesis 7	1.47916
UBE2U	ubiquitin-conjugating enzyme E2U (putative)	1.47816
CILP2	cartilage intermediate layer protein 2	1.47503
TFAP2E	transcription factor AP-2 epsilon (activating enhancer binding protein 2 epsilon)	1.47401
NHLRC2	NHL repeat containing 2	1.47226
POLR3H	polymerase (RNA) III (DNA directed) polypeptide H (22.9kD)	1.47159
ARHGAP12	Rho GTPase activating protein 12	1.46873
LOC285768	uncharacterized LOC285768	1.4653
LRRC3B	leucine rich repeat containing 3B	1.4605
SAP30BP	SAP30 binding protein	1.45864
CCDC137	coiled-coil domain containing 137	1.45469
C8ORF76	chromosome 8 open reading frame 76	1.44899
DHODH	dihydroorotate dehydrogenase (quinone)	1.44843
WDR75	WD repeat domain 75	1.44641
NOC2L	nucleolar complex associated 2 homolog (S. cerevisiae)	1.44532

TNFRSF18	tumor necrosis factor receptor superfamily, member 18	1.44218
GNL3	guanine nucleotide binding protein-like 3 (nucleolar)	1.44089
CCL16	chemokine (C-C motif) ligand 16	1.4402
SLC28A2	solute carrier family 28 (concentrative nucleoside transporter), member 2	1.43906
IL17B	interleukin 17B	1.43828
FOXE3	forkhead box E3	1.43681
CUTC	cutC copper transporter	1.43525
SLC25A22	solute carrier family 25 (mitochondrial carrier: glutamate), member 22	1.43518
HPR	haptoglobin-related protein	1.43476
DHX37	DEAH (Asp-Glu-Ala-His) box polypeptide 37	1.43075
RBP7	retinol binding protein 7, cellular	1.43013
RECQL4	RecQ protein-like 4	1.42619
SGTA	small glutamine-rich tetratricopeptide repeat (TPR)-containing, alpha	1.42056
EREG	epiregulin	1.41997
SRP68	signal recognition particle 68kDa	1.41782
CCT2	chaperonin containing TCP1, subunit 2 (beta)	1.41464
TRMT1	tRNA methyltransferase 1 homolog (S. cerevisiae)	1.41336
COA7	cytochrome c oxidase assembly factor 7 (putative)	1.4132
FGF19	fibroblast growth factor 19	1.41268
THUMPD3	THUMP domain containing 3	1.41018
THG1L	tRNA-histidine guanylyltransferase 1-like (S. cerevisiae)	1.40974
FOXN1	forkhead box N1	1.40789
SLC45A1	solute carrier family 45, member 1	1.40549
DHX30	DEAH (Asp-Glu-Ala-His) box helicase 30	1.39842
SUGT1	SGT1, suppressor of G2 allele of SKP1 (S. cerevisiae)	1.39672
C11ORF48		1.39638
PTPN11	protein tyrosine phosphatase, non-receptor type 11	1.39526
RBM45	RNA binding motif protein 45	1.3944
PPP3CA	protein phosphatase 3, catalytic subunit, alpha isozyme	1.39289
PTRH2	peptidyl-tRNA hydrolase 2	1.39261
PGGT1B	protein geranylgeranyltransferase type I, beta subunit	1.39186
SMNDC1	survival motor neuron domain containing 1	1.38806
ARHGEF19	Rho guanine nucleotide exchange factor (GEF) 19	1.38451
CYB5D1	cytochrome b5 domain containing 1	1.38372
TTI2	TELO2 interacting protein 2	1.38213

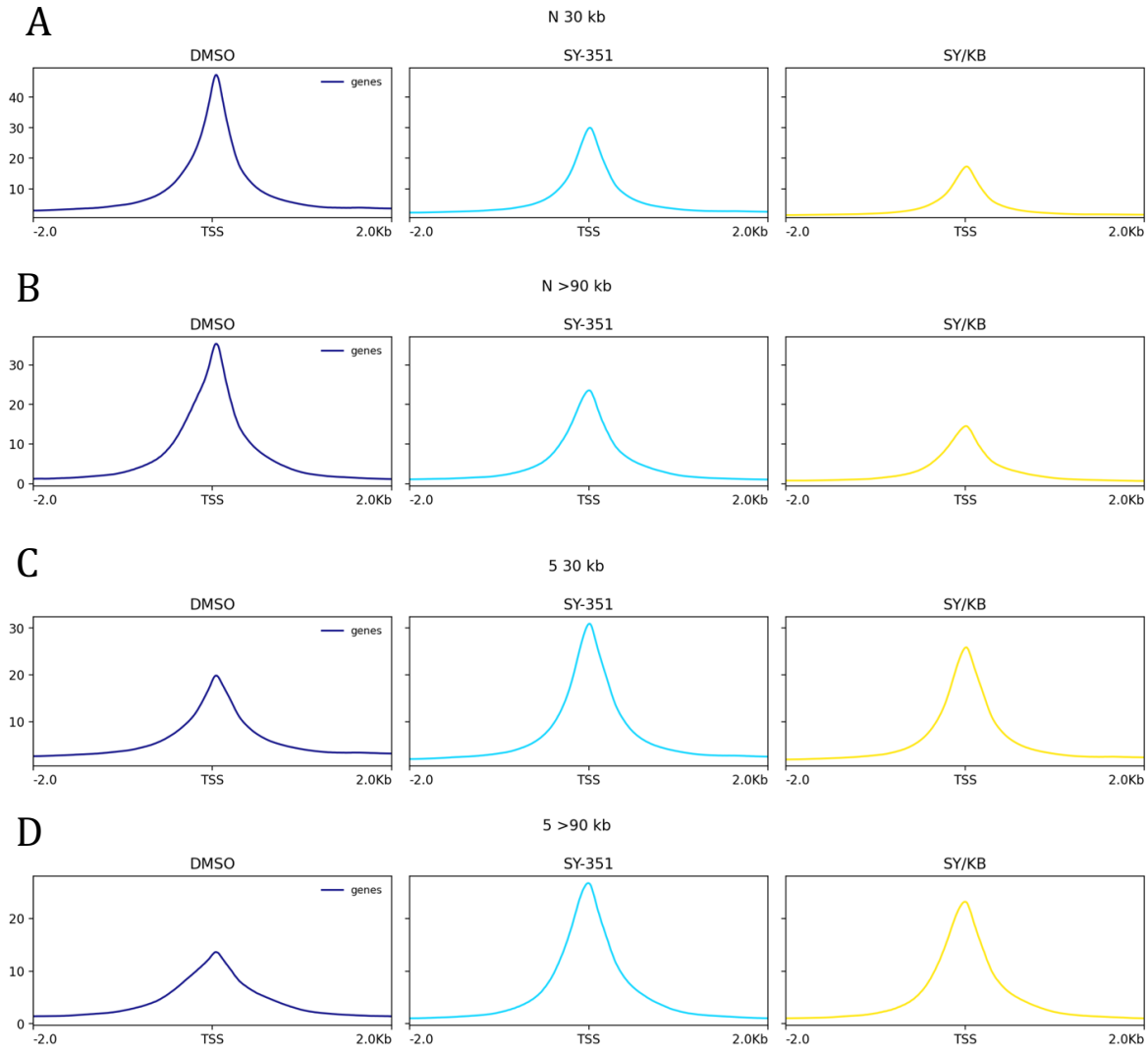
EMG1	EMG1 N1-specific pseudouridine methyltransferase	1.38137
PLRG1	pleiotropic regulator 1	1.37999
FAM136A	family with sequence similarity 136, member A	1.37946
NOVA2	neuro-oncological ventral antigen 2	1.37569
BRSK2	BR serine/threonine kinase 2	1.37432
EIF3A	eukaryotic translation initiation factor 3, subunit A	1.37253
DHX34	DEAH (Asp-Glu-Ala-His) box polypeptide 34	1.37192
MYH15	myosin, heavy chain 15	1.37083
MRPL27	mitochondrial ribosomal protein L27	1.36928
DPH2	DPH2 homolog (S. cerevisiae)	1.36604
WRAP73	WD repeat containing, antisense to TP73	1.36593
DHX9	DEAH (Asp-Glu-Ala-His) box helicase 9	1.36415
RPS6KB1	ribosomal protein S6 kinase, 70kDa, polypeptide 1	1.35696
SUPT16H	suppressor of Ty 16 homolog (S. cerevisiae)	1.35211
PMPCA	peptidase (mitochondrial processing) alpha	1.35092
TRAV17	T cell receptor alpha variable 17	1.35078
MIB2	mindbomb E3 ubiquitin protein ligase 2	1.35032
ZNF584	zinc finger protein 584	1.35017
FAM166B	family with sequence similarity 166, member B	1.34939
CLUH	clustered mitochondria (cluA/CLU1) homolog	1.34827
CHRNA10	cholinergic receptor, nicotinic, alpha 10 (neuronal)	1.3479

S30: Table Displaying List of 138 Highly Expressed Genes During IFN-γ Response[160,161]

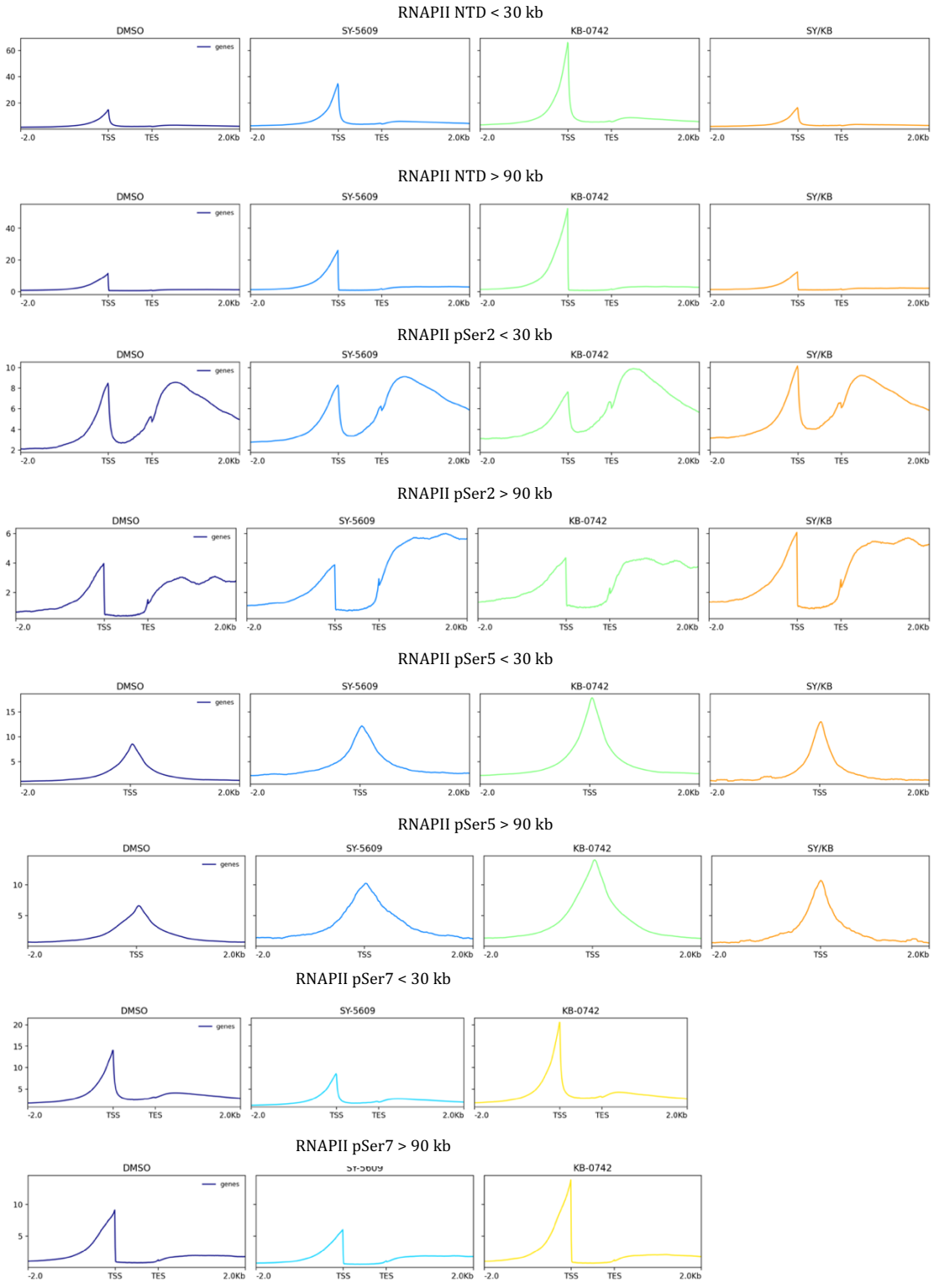
Gene List			
ADAR	NMI	STAT3	EGR1
ATP6V0B	PARP1	PARP9	CYTOR
B2M	PLOD2	UBE2L6	SIAE
BAG1	PLSCR1	HAPLN3	KDSR
BAK1	PMAIP1	PSMB9	DDX23
BBC3	PML	HELB	PARP14
BST2	PPP3CA	CSF1	HM13
BTG1	PPP5C	CCDC68	OSMR
C1S	PRAME	DTX3L	SPTBN1
CASP8	PSMB10	HK1	RUNX1
CEBPD	PSMB8	BCL6	ACTR3C
COL16A1	PSMB9	CA13	CCNL1
CSRP3	PSME1	FEZF1-AS1	ADAMTSL4-AS1
CYCS	RBBP4	LGALS3BP	SMG1
EIF2B1	RHOC	STAT2	EHD1
EIF2B1	SDCBP	SECTM1	CARS
ELK4	SEM1	ZC3HAV1	ERI1
EPS15	SF3A1	MSRB1	ENO1
FAS	SKP1	NUCB1	
FOSL1	SP110	RAB27A	
GBP1	SRP9	RTBDN	
HADH	SRSF2	GSTK1	
HADHB	SSBP1	S100A11	
HLA_A	STAT1	MIR4435-2HG	
HLA-C	TAP1	KIF2A	
HLA-E	TEAD4	IFNAR2	
ICAM1	TRIM21	TRIM25	
IFI16	TRIM26	C12ORF4	
IFI30	VAT1	PUS3	
IFI35	VEGFC	PLSCR4	
IFIT2	XRCC6	BCL3	
IFIT3	ZFP36L2	IL10RB	
IFITM1	ICAM4	C1QTNF6	
IL15RA	STEAP4	ALPK1	
IL6	NLRC5	COQ8B	
IRF1	REC8	FERMT2	
IRF9	APOL8	MAFF	
ISG15	SERPING1	GCLM	
MAP3K10	ETV7	CEP162	
NAP1L1	BATF2	SLC35F6	



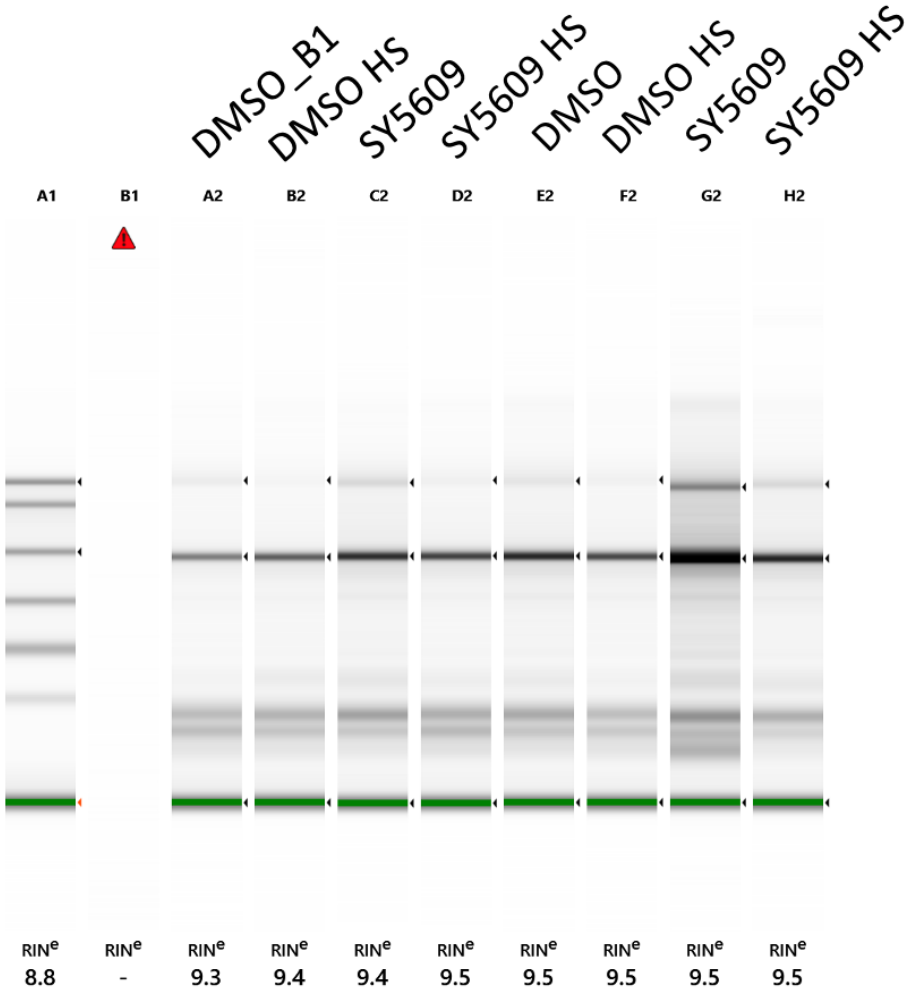
S31: SY-351 T1 hg38 genome metagene profiles by gene size. All reads were mapped to genes from the hg38 genome assembly. A, B correspond to RNAPII NTD. C,D correspond to RNAPII pSer5. A,C reflect enrichment at genes less than or equal to 30 kB in length. B,D reflect enrichment at genes greater than or equal to 90 kB in length. All samples were normalized to 1x read coverage prior to metagene analysis. Read duplications and blacklisted genomic regions were removed during analysis.



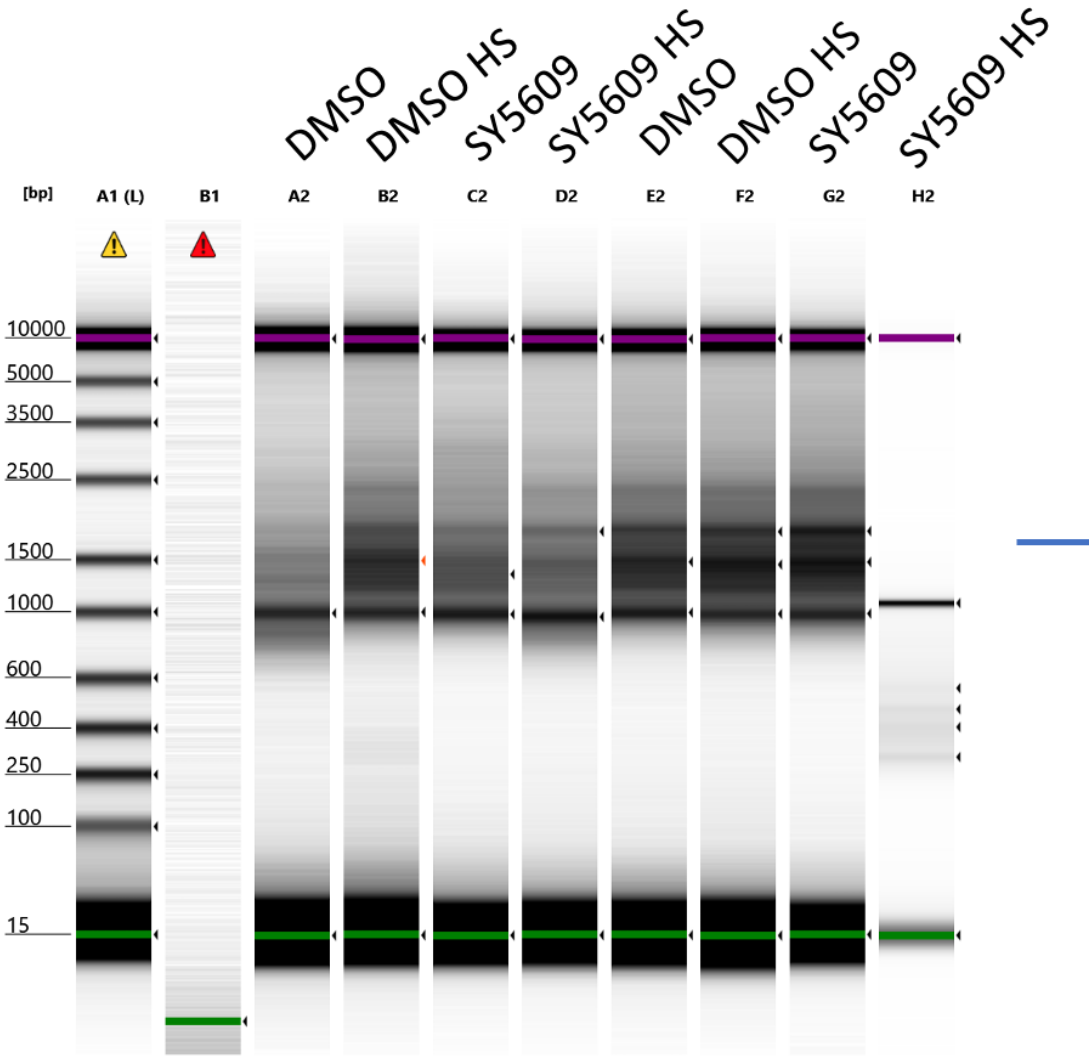
S32: SY-351 T2 hg38 genome metagene profiles by gene size. All reads were mapped to genes from the hg38 genome assembly. A, B correspond to RNAPII NTD. C,D correspond to RNAPII pSer5. A,C reflect enrichment at genes less than or equal to 30 kB in length. B,D reflect enrichment at genes greater than or equal to 90 kB in length. All samples were normalized to 1x read coverage prior to metagene analysis. Read duplications and blacklisted genomic regions were removed during analysis.



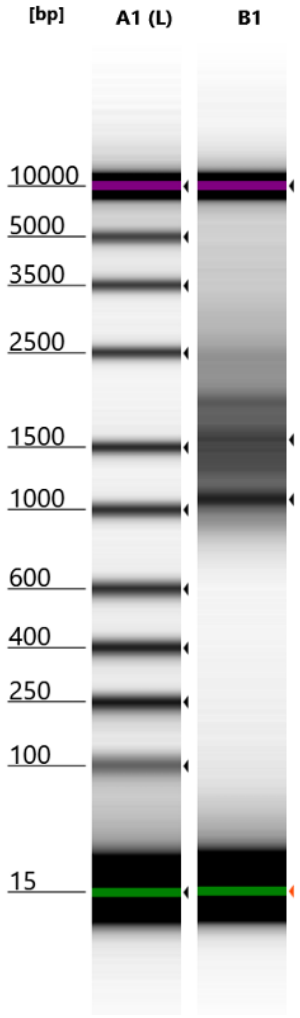
S33: SY-5609 B2 hg38 genome metagene profiles by gene size. All reads were mapped to genes from the hg38 genome assembly. All samples were normalized to 1x read coverage prior to metagene analysis. Read duplications and blacklisted genomic regions were removed during analysis.



S34: TapeStation Traces of Isolated RNA and Associated RIN scores. RNA was imaged using the Agilent High Sensitivity RNA ScreenTape Kit. 1 µL of each sample or D1000 Ladder was mixed with 3 µL Agilent D1000 Tapestation Buffer. Gels are imaged at roughly 60% saturation. RIN^e scores are equivalent to RIN scores produced by a Bioanalyzer.



S35: TapeStation Traces of Amplified cDNA. 1 µL of each sample or D5000 Ladder was mixed with 3 µL Agilent D5000 Tapestation Buffer. Gels are imaged at roughly 60% saturation.



S36: TapeStation Trace of Pooled cDNA Library for Long Read RNA Sequencing. 1 µL of each sample or D5000 Ladder was mixed with 3 µL Agilent D5000 Tapestation Buffer. Gels are imaged at roughly 60% saturation.

Table S37: RNA-Seq Reagents

Reagent	Supplier	Cat #
Cell Applications Inc MCDB 105 Medium	Fisher Scientific	50-608-877
Medium 199	Life Technologies	31153026
FBS	PEAK Serum	PS-FB3
Penicillin/Streptomycin	Invitrogen	15140
SY-5609	Syros Pharmaceuticals	Proprietary
TriZol LS	Invitrogen	10296028
RNeasy Mini Kit	Qiagen	74104
High Sensitivity RNA Screentape	Agilent	5067-5579
High Sensitivity RNA Ladder	Agilent	5067-5581
High Sensitivity RNA Sample Buffer	Agilent	5067-5580
Single Cell/Low Input cDNA Synthesis & Amplification Module	NEB	E6421S
Barcoded NEBNext single cell cDNA and Iso-Seq PCR primers	IDT	289807033-28980748
High Fidelity 2x PCR Mastermix	NEB	M0541S
Iso-Seq Express Oligo Kit	PacBio	101-737-500
D5000 DNA Screentape	Agilent	5067-5588
D5000 DNA Ladder	Agilent	5067-5590
D5000 DNA Sample Buffer	Agilent	5067-5589
SMRTBell Prep Kit 3.0	PacBio	102-182-700

Cell Culture:

OV90 cells (ATCC# CRL-11732) were grown to 70-90% confluency in a 1:1 mix of Medium 199 and MCDB 105, supplemented with sodium bicarbonate to a final concentration of 1.85 g/L, 15% FBS, and 1% Penicillin/Streptomycin. Cells were grown at 37°C and 5% CO_2.

Treatment Procedure:

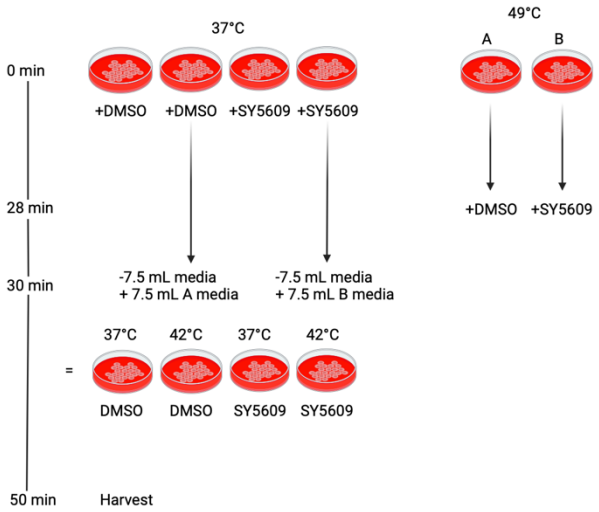


Figure S38: Treatment Scheme for Heat Shock Experiments. Identical flasks of cells (A, B) were incubated in a 49°C water bath for 30-40 minutes to generate 49°C pre-treated media. Cells for RNA-Seq were treated with SY5609 to a final concentration of 50 nM or an equivalent volume of DMSO and incubated at 37°C for 30 minutes. Just prior to the 30 minute timepoint, flasks in the 49°C waterbath were treated with either 50 nM SY-5609 or an equivalent volume of DMSO. At the 30 minute timepoint, half of the media from the 37°C incubated cells was removed, and an equivalent volume of 49°C media was added, effectively bringing the temperature of the media on cells for RNA-Seq to 42°C. This generated heat shock conditions for instantaneously. These cells were then placed into a 42°C water bath for 20 minutes prior to harvest. Non-heat shocked cells for each inhibitor condition remained at 37°C. Two biological replicates were generated. Cells were harvested in ice cold PBS and immediately resuspended in 400 uL of Trizol LS. Cells were stored at -20°C for future use. Figure created in BioRender.

RNA Isolation:

Samples stored in TriZol were thawed, and 300 uL was transferred to a new tube allowed to sit for five minutes at room temperature to ensure complete cell lysis and RNA isolation. After this, 80 uL of chloroform was added, tubes were gently mixed, and incubated at room temperature for 2-3 minutes. Phases were separated via centrifugation and the aqueous RNA phase was transferred to a fresh tube. 1.5 uL of GlycoBlue and an equivalent sample volume of isopropanol was added to the RNA. Tubes were incubated at -20°C for 30 minutes, and RNA was pelleted via centrifugation at 12,000 x g for ten minutes at 4°C. Supernatant was removed, and pellets were washed in 300 uL 75% ethanol. Ethanol was removed, and pellets were dried and resuspended in 50 uL RNase-free water.

DNA digestion was performed using the Qiagen RNeasy Mini Kit. Resuspended RNA was incubated with 2.5 uL of DNase I at room temperature for ten minutes. RNA was re-isolated using RNeasy spin columns and eluted using 50 uL of RNase-free water. RNA was quantified via Qubit and quality was assessed using the Agilent 2200 Tapestation and corresponding High Sensitivity RNA Tapestation reagents. RNA was too dilute for library preparation, so 47 uL of RNA sample was mixed with 1.5 uL of GlycoBlue, 20.8 uL of 1 M NaOAc, and 208 uL 100% ethanol. RNA precipitated overnight at -76°C, and was re-pelleted and isolated according to the above protocol. RNA was quantified and qualitatively assessed with the Qubit and Tapestation, respectively (Figure S35).

Library Preparation:

Libraries were prepared with the PacBio SMRTBell 3.0 IsoSeq Prep Kit according to the kit instructions. 228 ng of RNA from each experimental condition was reverse transcribed, and subjected to a 1.3X bead based cleanup. The resulting cDNA was PCR amplified for 12 cycles. cDNA amplification utilized uniquely barcoded primers sourced from IDT, in accordance with Appendix 3 of the PacBio kit. cDNA was isolated using a bead based cleanup and quantified via Qubit. Due to an error in cDNA quantification, quantity was underestimated, and cDNA was subjected to an additional 2 cycles of PCR amplification

and an additional bead based cleanup. cDNA was then (correctly) quantified via Qubit and visualized on an Agilent 2200 Tapestation using D5000 Tapestation reagents (Figure S36). 62.5 ng of each cDNA library was used to create a library pool with 500 ng of total cDNA in 46 uL. The pooled libraries were subjected to end repair and A-tailing, followed by SMRTBell adapter ligation. Adapter dimers and contaminants were removed via a 1.3X bead based cleanup, and a nuclease treatment was used to remove damaged or non-circularized SMRTBell templates. Nucleases were inactivated via a 15 minute incubation at 37°C and removed via a 1.3X bead based cleanup. The library was quantified via Qubit and again visualized on the Tapestation using D5000 reagents (Figure S37). Samples were sequenced by the Ramani Lab at UCSF. Data analysis is also being performed by the Ramani lab and is currently in progress.

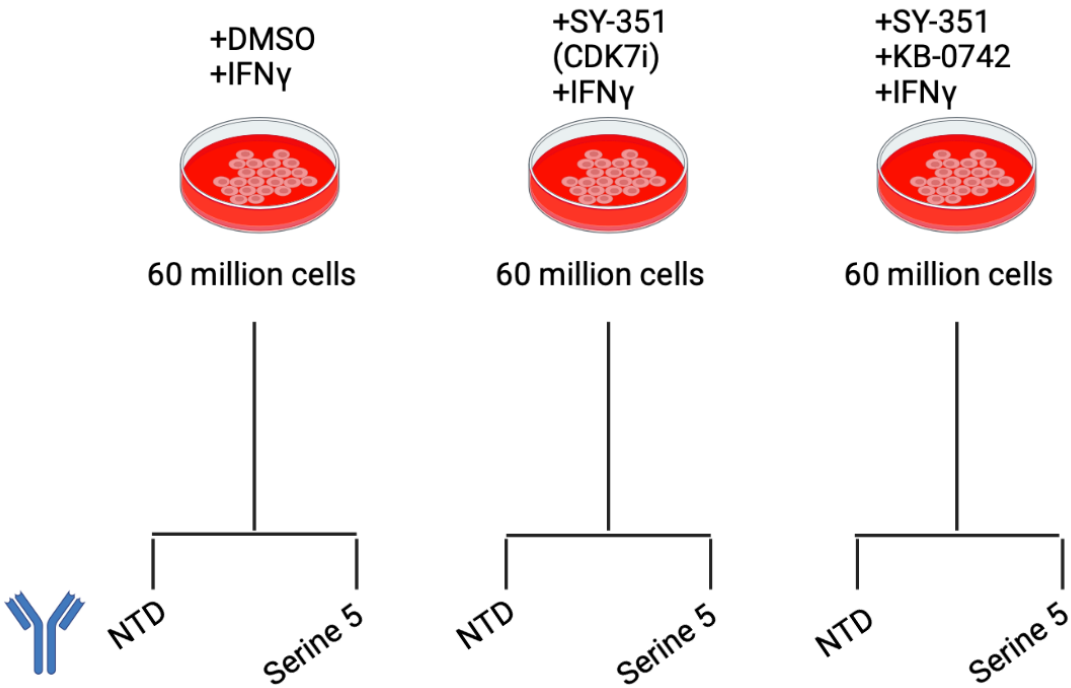


Figure S39: Schematic of SY-351 ChIP-Seq experimental conditions. Cells were treated with DMSO, SY-351, or SY-351 and KB-0742 to produce two technical replicates of ~60 million cells/condition. Each treatment group was subdivided into four IPs. All cells were treated with IFNγ. Created in Biorender. ChIP procedure is identical to that described in Methods section.