Title
Elucidating the Role of Foggy, a Transcription Elongation Factor, in Zebrafish Development

Author
Krishnan, Keerthi

Publication Date
2008-01-03

Peer reviewed|Thesis/dissertation

Elucidating the Role of Foggy, a Transcription Elongation Factor, in Zebrafish Development

by

Keerthi Krishnan

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

[redacted]

UMI Number: 3289344

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

I would like to thank Dr. Su Guo for being my Ph.D advisor for the past 4.5 years. I will always remember the experiences and the lessons I have learned in her lab. I would like to thank Nathan Salomonis for guiding me expertly through the maze of computational requirements of my Thesis project. I would like to thank Drs. Rubenstein and Ingraham for serving on my Thesis Committee and for their guidance and support. I would also like to thank Drs. Jan, Margeta, Guy, and Szoka for their advice and valuable insights during my graduate training. I would like to acknowledge the UCSF community for offering me help and guidance on wide-ranging issues from technical experimental expertise to career development.

In addition to research experiences, my life has been enriched by the caliber of people that I have met at UCSF. Some of these people have become my best friends and have forever changed my perspectives on life and science. On that note, I would like to thank Dr. Thomas Mueller and Billy Lau for being my friend and guide for the past years. They made my graduate school experience sparkle with excitement at times and plunged me into depths of introspection at other times and for this interaction, I will forever be grateful. In addition, my special thanks go to Billy Lau for always being there when I needed him. Obtaining my Ph.D degree would have been a lot harder and much less fun without his constant goading and reminders to work diligently.

Finally, I would like to thank my sister, Preethi for providing me encouragement and my parents for instilling the value of a good education and the strength to pursue it.

Abstract

Spt5 (also known as DSIF-P160, Tat-CT, and Foggy) is an evolutionarily conserved essential protein that represses as well as stimulates transcription elongation *in vitro*. Although less than a handful of genes are shown to be under direct regulation by Spt5, the prevalence and identity of its target genes is poorly defined *in vivo*.

Through expression profiling studies on ~10,000 protein-coding genes, a surprisingly small fraction (less than 5%) was differentially expressed between *fog*^{sk8} mutants and their wild-type (WT) siblings while less than 1.5% was differentially expressed between *foggy* ^{m806} and their WT siblings. Further classification analyses revealed that these differentially expressed genes are involved in diverse biological pathways ranging from pattern specification to stress response. Many previously uncharacterized, novel genes were also differentially expressed in these mutants.

A permutation based analysis revealed that genes upregulated in *fog*^{sk8} and *foggy*^{m806} embryos have significantly shorter gene length compared to the average gene length in zebrafish, suggesting an additional repressive level of regulation of shorter genes by Foggy/Spt5. *In vivo* chromatin immunoprecipitation further uncovered genes that are directly bound and regulated by Foggy/Spt5 at the level of transcription elongation. Furthermore, Foggy and RNAPII occupancies suggest the poised transcriptional status of many upregulated genes, and active transcription of many downregulated genes in WT embryos.

Our comparison of RNAPII occupancy between WT and *sk8* embryos identified Growth Arrest and DNA-damage inducible gene (*gadd45b*) to be a novel target for

repression by Foggy. Transcription elongation control of such inducible genes might be crucial to integrate signals from multiple pathways to coordinate regulation of important signaling pathways that ultimately determine cell fate and survival. In conclusion, our results identify a small but functionally diverse set of genes that are regulated by Foggy/Spt5 and provides interesting insights into the identities and characteristics of target genes of Foggy/Spt5 in vertebrate development.

Table of Contents

List of Tables

List of Figures

I. Introduction

Model for Gene Regulation

The traditional view of gene expression considered the process of transcriptional activation as distinct, consecutive and linear steps encompassing chromatin remodeling, synthesis (initiation, elongation, termination) of mRNA, post-transcriptional processing of the primary transcript (5'capping, 3' polyadenylation and splicing), export of mRNAs to the cytoplasm for translation (Orphanides and Reinberg 2002). Most of our knowledge of each of the steps in this pathway from gene to protein was generated by reductionist approaches (e.g., classical biochemistry). Although this approach has illuminated the identities and functions of individual proteins (e.g., chromatin-remodeling factors, DNA-binding transcription factors) by providing a mechanistic view of each individual step in the process, the continuity between the steps and the potential dynamics and non-linearity among the processes in living cells cannot be revealed by this approach alone.

In recent years, genetic studies in diverse model organisms from yeast to zebrafish, large-scale mapping of protein interaction networks, genome-wide molecular studies (differential transcript expression, protein occupancy on chromatin) have suggested that the steps involved in regulating gene expression are physically and functionally connected and represent subdivisions of a continuous process with significant interdependency between steps (Fig.1). Many of the factors essential for gene regulation have multiple roles and function in large complexes, merging the previously known individual steps as a continuous process. This contemporary view of gene expression (Orphanides and Reinberg 2002) provides us with a unified theory of gene

regulation *in vivo* and highlights the complexities involved in gene regulation. Many of the interdependent steps occur during transcription elongation stage, which is the focus of this work.

Transcription

Transcription of mRNA is divided into three stages (initiation, elongation, termination). Transcription initiation involves binding of transcription factors to their binding sites on the DNA, recruitment of RNA Polymerase II (RNAPII) and associated factors to bind to promoter sites and RNA synthesis commences. Transcription of mRNA at the initiation step is extensively studied and viewed as an important mode of regulating gene expression. The next step, transcription elongation, was considered as a trivial addition of ribonucleoside triphosphates to the growing mRNA chain for many years. It is now apparent in the contemporary view of gene regulation that transcription elongation is a critical stage that couples mRNA synthesis to processing of the transcript (5' capping, splicing, 3' polyadenylation) on the chromatin, ensuring efficiency and accuracy of transcription (Saunders et al. 2006; Sims et al. 2004). Capping of the 5' end of the RNA is the first of these modifications and occurs as soon as the 5' end of the nascent transcript becomes accessible, usually after transcription of 20-40 nt of RNA. It was known that RNAPII complex entered a pausing state shortly after clearing the promoter (Saunders et al. 2006). It is possible that this pausing occurs to allow time for 5' capping, and that RNAPII will not continue until this protective modification has been added to the nascent mRNA. This is analogous to the "checkpoints" that operate during

the cell cycle to ensure that each phase of the cycle is complete before the next begins. Alternatively, it is possible that the pausing allows time for the necessary transition of

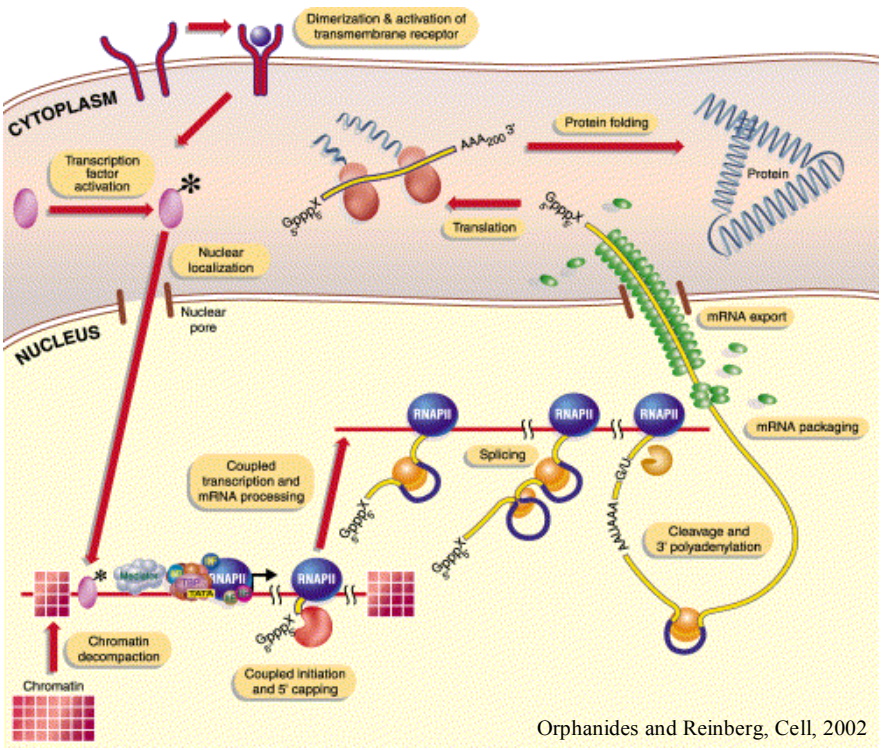


Figure 1: The unified theory of gene regulation suggests that steps involved in regulating gene expression are physically and functionally connected (by multiple functions of proteins) and represent subdivisions of a continuous process with significant interdependency between steps to ensure accuracy and efficiency of transcription. Many of the interdependent steps occur during transcription elongation where mRNA synthesis is coupled to processing, highlighting the importance of regulating transcription elongation.

different proteins (e.g., specific elongation factors, polyadenylation factors, and splicing factors) onto the RNAPII complex. It is evident that transcription elongation factors influence splicing and that splicing factors increase transcriptional elongation (Fong and

Zhou 2001). For example, introns are necessary for efficient RNAPII transcription (Furger et al. 2002), mutations in elongation factors cause splicing defects (accumulation of unspliced RNA) or nuclear degradation of unspliced pre-mRNAs in yeast (Kornblihtt et al. 2004).

In addition to these general processing functions required for most mRNAs, transcription elongation has also been studied as a specific and regulated step of gene expression for a handful of genes such as heat shock genes (*hsp70*), proto-oncogenes (*c-fos*), and viral genes (Chiang et al.) (Saunders et al. 2006). In the case of *hsp70* gene transcription in *Drosophila*, RNAPII, elongation factors and general transcription machinery is paused at the 5' end of the gene, under non-inducing conditions. Within few minutes after heat shock stimuli, RNAPII complex is relieved from its paused state and transcribes mRNA immediately, allowing for the fast production of the *hsp70*-encoding chaperone protein to cope with the stress (Andrulis et al. 2000; Giardina and Lis 1993; Li et al. 1996; O'Brien and Lis 1991; Rasmussen and Lis 1993; Rougvie and Lis 1988; Wu et al. 2005; Wu et al. 2003). Transcription elongation factors play repressive (maintaining RNAPII pause) and stimulatory (increasing RNAPII processivity) roles along *hsp70* gene.

In the case of *c-fos* (Fivaz et al. 2000; Pinaud and Mirkovitch 1998; Plet et al. 1995; Ryser et al. 2007; van Haasteren et al. 2000; Yamada et al. 2006) and *hiv* (Bourgeois et al. 2002; Lanchy et al. 1996; Wu-Baer et al. 1998), transcription elongation factors play a stimulatory role in proper transcription of these genes. Rapid gene transcription in response to stimuli is a recurring theme during many biological processes like development and hence, it is plausible that regulation of transcription elongation

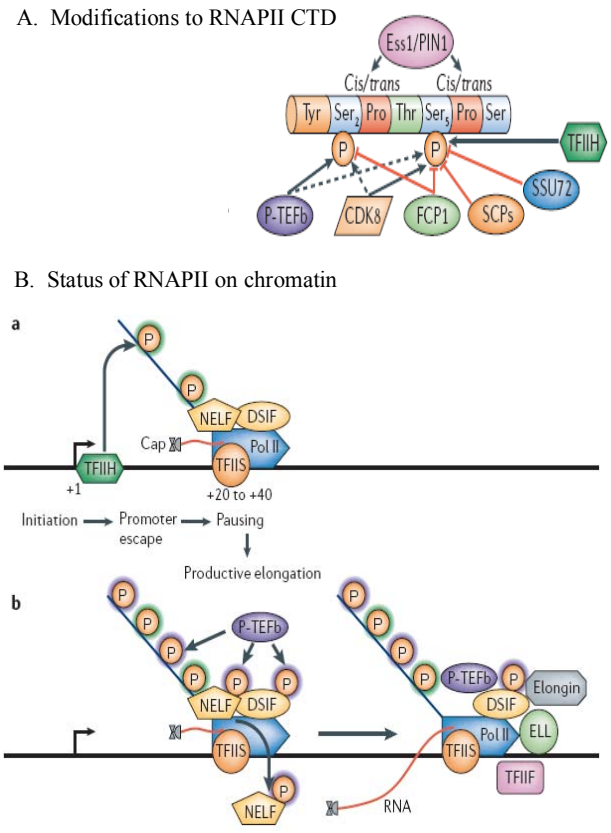
might be more prevalent than once considered. Indeed, recent genome-wide analyses also find significant RNAPII occupancy at the 5' end of many genes, whether or not their transcripts can be detected (Guenther et al. 2007; Kim et al. 2005; Mikkelsen et al. 2007).

Status of RNAPII complex during transcription cycle

RNAPII complex, the engine that drives mRNA transcription, consists of 12 subunits that are highly conserved throughout eukaryotes. The largest subunit Rpb1 contains C-terminal domain (CTD), which undergoes modifications throughout the transcription cycle (Fig. 2A). The RNAPII CTD consists of multiple repeats of the heptapeptide sequence YSPTSPS. Though the reason is unclear, it is interesting to note that the number of these repeats increases with genomic complexity: 26 in yeast, 32 in *C.elegans*, 45 in *Drosophila*, and 52 in mammals (Sims et al. 2004). The CTD is a 'loading dock' for multiple proteins that bind RNAPII. The identities and functions of the factors that bind to the CTD vary depending on the position of RNAPII on the chromatin. General transcription factors bind CTD at the promoter, while polyadenylation factors and splicing factors bind CTD away from the promoter (Kim et al. 2004). How do different factors bind to the same CTD? The answer lies in the multiple modifications such as phosphorylation of specific residues and *cis-trans* conformation changes of repeats, which allow the CTD to undergo three-dimensional structural changes.

Studies using different antibodies to RNAPII have shown that phosphorylation modifications to the CTD can be correlated to the status of the RNAPII on the chromatin and to the occupancy of different factors on the CTD; the so-called 'CTD code'. When

RNAPII is present on the promoter, its CTD is largely unphosphorylated, and when it is involved in productive elongation, the CTD is hyperphosphorylated (Fig. 2B). Initial studies with specific antibodies distinguished that CTD phosphorylation of Serine 5 residue correlated with transcription initiation and early elongation, while phosphorylation of Serine 2 correlated with RNAPII that has traveled farther from



Saunders and Lis, Nature reviews molecular cell biology, 2006

Figure 2: (A). Modifications to CTD repeats (YSPTSPS) by different kinases, phosphatases and isomerases allow many different factors to bind and function during transcription. (B). Status of RNAPII on chromatin with respect to the stages of transcription. (a). RNAPII complex is held in paused status after promoter clearance by NELF and DSIF. (b). P-TEFb, a kinase, phosphorylates CTD and DSIF to remove pausing and continue elongation of nascent mRNA. Multiple other elongation factors associate with RNAPII during productive elongation.

the promoter. It was found later that the antibody recognizing phosphorylated -Ser 2 also recognized phosphorylated-Ser 5 (Jones et al. 2004). Nevertheless, the general trend remains true; the status of RNAPII complex and its associated factors can be determined by using different antibodies against the CTD and the location of the RNAPII complex along the length of the gene. RNAPII CTD goes through cycles of phosphorylation and de-phosphorylation by multiple highly conserved kinases and phosphatases (Saunders et al. 2006). It is unclear how and when different kinases are recruited to the RNAPII CTD located at specific gene loci.

Several amino acids of the RNAPII CTD also undergo *cis-trans* conformation changes by virtue of their association with ESS1, a protein that functions as a peptidyl-prolyl-isomerase (Morris et al. 1999; Myers et al. 2001; Shaw 2007; Wilcox et al. 2004; Wu et al. 2003; Wu et al. 2000) (Fig. 2A). This modification adds another level of structural change to the CTD, allowing multiple different factors to dock and un-dock from the CTD. RNAPII complex can also be ubiquitylated on the chromatin during transcriptional arrest or as a natural process of removing stalled RNAPII complex from the chromatin (Somesh et al. 2005). Ser-5 phosphorylated CTD is averse to ubiquitylation compared to Ser-2 phosphorylated CTD, suggesting that elongating RNAPII complex might be subjected to degradation. It is also possible that the ubiquitylation provides another function besides degradation, similar to ubiquitylation of histones (Somesh et al. 2005).

Transcription elongation factors

A transcription elongation factor was defined as any molecule that affects the activities of or is associated with the RNAPII transcription elongation complex (TEC) (Sims et al. 2004). Transcription elongation factors are highly conserved across species and were initially identified through genetic studies in yeast. So far, 21 elongation factors, several with multiple subunits, have been identified (Sims et al.). Transcription elongation factors are ubiquitously expressed in many organisms. Many of the elongation factors do not affect the rate at which the elongating RNAPII complex traverses the gene (elongation rate) or the ability of the elongating RNAPII to travel the entire length of the gene (processivity) (Mason and Struhl 2005). Instead, many of these factors might affect chromatin status or modulate mRNA processing events during the elongation stage of transcription. Thus, in *in vitro* biochemical context (where chromatin status is not measured), there is a high degree of redundancy among elongation factors, and the physiological relevance of these factors in an *in vivo* context is largely unknown. The focus of this study is to elucidate the role of Suppressor of Ty 5 (SPT5) using zebrafish as a model organism.

Discovery of SPT5

In *S. cerevisiae*, the *SPT* genes (Suppressor of Ty) were identified by selection for extragenic suppressors of transcriptional defects caused by Ty insertions (Hartzog et al. 1998; Swanson et al. 1991; Swanson and Winston 1992). Seventeen *SPT* genes were identified by this process and based on genetic and molecular analysis categorized into three groups. SPT5, the subject of this study, along with SPT4 and SPT6 were grouped

together with histone proteins H2A and H2B. The categorization of SPT5 with histone proteins led to the initial hypotheses that these proteins might be important for establishment or maintenance of chromatin structure. Further genetic analysis revealed that mutations in SPT4, SPT5 and SPT6 suppress the defect in Ty transcription caused by mutations in *snf* genes (Happel et al. 1991), suggesting that the SPT4-SPT5-SPT6 complex would be required for repression of gene expression. SPT5 and SPT6, not SPT4, are essential for mitotic growth.

Biochemical characterization of SPT5

The search for factors that inhibit transcription elongation in the presence of the nucleotide analog 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) (Chodosh et al. 1989) in crude HeLa cell extracts led to the purification of Spt4 and Spt5 as a tight complex named DSIF (DRB-sensitivity-inducing factor) (Wada et al. 1998). When purified, recombinant DSIF was added to partially reconstituted transcription system containing general transcription factors and highly purified RNAPII, the synthesis of a 100-nucleotide (nt) transcript from an artificial G-free cassette was inhibited by increasing amounts of DSIF, suggesting a negative role in elongation (Fig. 2B). DSIF was also able to stimulate elongation of ~400 nt artificial transcripts under conditions of limiting concentrations of ribonucleoside triphosphates. Without DSIF, the rate of RNA chain elongation was very slow and short-length RNA transcripts of ~30 nt accumulated in the reaction, suggesting that DSIF stimulated the rate of accumulation of full length transcripts.

It is now clear that the partially purified factors used in the DRB sensitivity assay to identify DSIF contained another complex, Negative Elongation Factor (NELF), which acts in concert with DSIF to inhibit elongation. In the absence of DSIF, the NELF complex is unable to bind and repress RNAPII, showing that DSIF is required to inhibit transcription elongation *in vitro* (Yamaguchi et al. 2002). Productive elongation is resumed when P-TEFb (Positive Transcription Elongation Factor) phosphorylates RNAPII and Spt5, thereby dissociating NELF from the RNAPII TEC.

DSIF physically associates with multiple components of the transcriptional machinery (TFIIF, TFIIS, CSB, Spt6, FACT, Chd1, Paf complex, Capping enzyme, capping methyltransferase and nuclear exosome) (Sims et al. 2004), suggesting roles in general mRNA processing events related to its stimulatory activity. One example between SPT5 and the capping enzyme (CE) highlights the function of SPT5 in mRNA processing events. CE is recruited to the transcription complex by its interaction with DSIF and Ser-5 phosphorylated CTD. This recruitment of CE to the transcription complex disables NELF-mediated repression of transcription (Mandal et al. 2004) and the purpose of the SPT5/CE interaction is to ensure timely capping of the nascent pre-mRNA before committing RNAPII to processive elongation (Pei and Shuman 2002). In yeast, mutations in *spt4* and *spt5* cause splicing defects (accumulation of unspliced RNA) or nuclear degradation of unspliced pre-mRNAs (Xiao et al. 2005). This suggests the Spt5 is important for proper splicing of mRNA and disposing of improper mRNAs, thereby playing roles in mRNA stability and surveillance.

Furthermore, DSIF has been shown to affect RNAPII processivity (the ability of the elongating RNAPII to travel the entire length of the gene), suggesting the importance

of DSIF to maintain the stability of the RNAPII TEC on the chromatin template in yeast (Mason and Struhl 2005).

Protein Domains of SPT5

SPT5 is a highly conserved, nuclear protein that is present in a variety of species from yeast to humans (Fig. 3). Interestingly, evolutionary analysis of the SPT5 protein sequence and domains have led to the conclusion that SPT5 was present in the last common ancestor of the three kingdoms of cellular life, archaea, eubacteria and eukarya (Ponting 2002), suggesting that this might be one of the earliest transcription elongation factors. NusG, an anti-terminator protein in bacteria is highly homologous to SPT5. In zebrafish, Spt5/Foggy is a 1084 amino acid (aa) with many distinctive structural features. The N-terminal region is highly acidic, contains putative nuclear localization signal and might function in modulating chromatin structure. The central region contains KOW motifs, which function as RNA-binding motifs (Kyrpides et al. 1996). There seems to be a disagreement in the number of KOW motifs in Spt5 protein (Hartzog et al. 1998; Yamaguchi et al. 1999). Bioinformatic analysis determined that there are 5 KOW motifs in yeast and 6 in mammals (Ponting 2002). Interestingly, the sixth KOW motif in mammals is found in the C-terminus region.

The C-terminus region contains two repeat regions (CTR1, CTR2) and a non-repeat region, which is unique to SPT5 protein in multi-cellular organisms. CTR1 contains several heptad repeats that are phosphorylated by P-TEFb, and is critical for DRB-mediated repression (Ivanov et al. 2000). The CTR2 region contains multiple serine and threonine repeats and may serve as potential sites for phosphorylation.

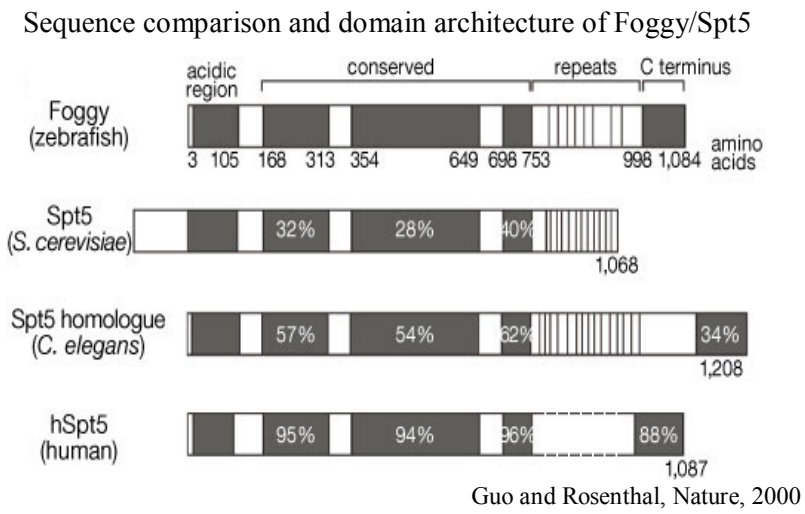


Figure 3. The domain structure analysis of Foggy/Spt5: The KOW domains are present in the central conserved domain. The repeats are sub-divided into CTR1 and CTR2. The C-terminus region is found in multi-cellular organisms and not found in yeast. The evolutionary sequence comparison of Spt5/Foggy shows regions of high similarity between zebrafish Foggy and Spt5 in other species.

Deletion analysis of Spt5 was used to identify structural features required to bind Spt4 (minimal domain of 176-270 aa) and RNAPII (313-420 aa). The N-terminal region, the Spt4-binding domain, the KOW domains and CTR1 are important for DSIF activity, highlighting the modular properties of this protein. It has been shown that in the absence of the CTR1 domain, SPT5 can bind to RNAPII but is not able to modulate its processivity (Ivanov et al. 2000). SPT5 is also methylated by protein arginine methyltransferases (PRMTs) in a region between KOW domains 3 and 4, which regulates its interaction with RNAPII. In addition to phosphorylation by P-TEFb, methylation of SPT5 by PRMTs results in stimulating elongation (Kwak et al. 2003).

Function of SPT5 during development

In vivo studies in worms, flies and zebrafish have provided considerable evidence for the physiological importance of Spt5 in multi-cellular organisms. Spt5 transcripts and protein is maternally deposited and Spt5's function is essential for normal development. Inactivation of Spt5 through RNAi in *C. elegans* arrested embryonic development prior to gastrulation, but many differentiated cell types were formed (Shim et al. 2002), suggesting that not all transcription was blocked. The expression of *hsp70*, a known target of transcription elongation, was induced in *spt-5(RNAi)* embryos under heat-shock conditions suggesting that heat shock genes can be transcribed without Spt5 in worms.

In *Drosophila*, a missense mutation (W049) altered the C-terminal domain of Spt5, and displayed locus-specific effects on the expression of segmental patterning genes (Jennings et al. 2004). Specifically, the expression patterns of gap genes were similar to wild-type patterns, but striped patterning of the primary pair-rule genes, *even-skipped* and *runt* was disrupted. The striped expression pattern of *even-skipped* was broadened in the mutant embryos, indicating that Spt5 is likely to be a direct, negative regulator of this target gene. By contrast, heat shock gene expression upon heat shock condition was reduced in the mutants, suggesting that Spt5 acts both positively and negatively on transcription *in vivo* depending on context.

A zebrafish mutation, fog^{m806} allele, altered one amino acid (V1012D) in the C-terminus of Spt5 protein, resulting in distinct defects in neuronal specification. These mutants display decreased dopamine neurons and a corresponding increase in serotonin neurons in the developing forebrain (Guo et al. 2000). *In vitro* assays measuring

repressive activity (DRB-sensitivity assay) and stimulatory activity (elongation under limiting nucleotide concentrations) suggests that the mutant protein had lost its repressive activity, while maintaining its stimulatory activity. Together, these results suggest that the repressive activity of Spt5/Foggy is required for proper specification of neurons in the developing zebrafish.

Another hypomorphic allele, *fog* ^{fh20}, displayed defects in posterior migration of facial branchiomotor neurons from rhombomere 4 (r4) into r6 and r7 of the hindbrain (Cooper et al. 2005). When cells from the mutant embryos were transplanted into a wild-type host, the mutant facial branchiomotor neurons survive to at least 5 days post fertilization, while failing to migrate posteriorly. These results suggested that SPT5 is important for transcribing genes related to migration cell-autonomously.

Severe truncation (the *fog*^{sk8} allele) or deletion (the *fog*^{s30} allele) of the Spt5 product resulted in broad deficits in embryonic development (Keegan et al. 2002). The *sk8* and *s30* mutants displayed developmental retardation by 20 hours post fertilization (hpf) followed by further deterioration and death by 48 hpf. The *fog*^{sk8} embryos showed several gross abnormalities including pigmentation, tail, ear and cardiac defects. Further analysis of these mutant embryos showed that Foggy/Spt5 is required for normal differentiation of the precardiac mesoderm into myocardial tissue. The expression of *hsp70* was reduced in mutant embryos under heat shock stimulus, suggesting that Spt5 is required for *hsp70* induction in zebrafish.

These *in vivo* mutational studies in multiple organisms suggest the existence of specific target genes of Spt5 which function in many diverse biological processes.

Goal of this study

In spite of the initial characterization of the *in vivo* physiological roles of Foggy/Spt5 in different model organisms, many questions remain. A).What genes are regulated by Foggy/Spt5 at the level of transcription elongation? B).What type of elements (*cis* regulatory sequences, gene structure characteristics, function of the genes) defines Foggy/Spt5's target genes?

We have approached these questions by using multi-faceted approaches with *fog*^{sk8} and in *foggy* ^{m806} mutant embryos. Chapter two elaborates on our interesting findings from *fog*^{sk8} mutants and chapter three summarizes our findings from *foggy* ^{m806} mutant analysis. Chapter 4 provides conclusions and future directions based on these studies.

II. Identification of *in vivo* targets of Foggy/Spt5, a dual regulator of transcription elongation, during zebrafish development

Keerthi Krishnan, Nathan Salomonis*, and Su Guo

Department of Biopharmaceutical Sciences, Programs in Human Genetics and Pharmaceutical Sciences and Pharmacogenomics, University of California, San Francisco, California, 94143-2811

*Gladstone Institute of Cardiovascular Research,

Nathan Salomonis performed the computational analysis for this study

Manuscript in preparation for publication

Introduction:

The traditional view of gene expression contained distinct, consecutive, linear steps encompassing chromatin de-condensation, transcription initiation by DNA-binding factors, recruitments of the general transcription machinery, elongation of the nascent mRNA, post-transcriptional processing, and splicing of the mature transcript (Orphanides and Reinberg 2002). In recent years, genetic studies in diverse model organisms, conventional biochemical approaches, large-scale mapping of protein interaction networks, genome-wide differential transcript expression and protein occupancy studies have suggested that the steps involved in regulating gene expression are physically and functionally connected and represent subdivisions of a continuous process with significant interdependency between steps (Orphanides and Reinberg 2002). From these studies, transcription elongation has emerged as a critical stage that couples mRNA synthesis to processing of the transcript (5' capping, splicing, 3' polyadenylation) on the chromatin, ensuring efficiency and accuracy of transcription (Saunders et al. 2006; Sims et al. 2004). In addition to these general processing functions required for most mRNA, transcript elongation has also been studied as a specific and regulated step of gene expression for a handful of genes such as heat shock genes (*hsp70*), proto-oncogenes (*c-fos*), and viral genes (Chiang et al.) (Saunders et al. 2006). The prevailing notion for this type of regulation is that it allows rapid inducibility in response to stimuli. The quest to identify the underlying commonalities among the genes repressed at the elongation level at the *cis* regulatory level has not yet yielded promising results, possibly due to a small number of target genes identified thus far.

A transcription elongation factor has been defined as any molecule that affects the activities of or is associated with the RNA Polymerase II (RNAPII) transcription elongation complex (TEC) (Sims et al. 2004). Transcription elongation factors identified to date are highly conserved across species and were initially identified through genetic studies in yeast followed by biochemical characterization in *in vitro* systems. One such elongation factor, Suppressor of Ty 5 (Spt5) was the focus of this study. Spt5, together with Spt4 and Spt6, was initially discovered by genetic studies in *Saccharomyces cerevisiae* (Winston et al. 1984), and later found to be essential for transcription via modulation of the chromatin structure (Bortvin and Winston 1996; Swanson and Winston 1992). The search for factors that inhibit transcription elongation in the presence of the nucleotide analog DRB in crude mammalian extracts led to the purification of Spt4 and Spt5 as a tight complex named DSIF (DRB-sensitivity-inducing factor) (Wada et al. 1998).

Over the years, Spt5 has been reported to have dual properties of repressing and stimulating transcription elongation in different contexts (Bourgeois et al. 2002; Wu-Baer et al. 1998; Yamaguchi et al. 1999). It is now clear that the partially purified factors used in the DRB sensitivity assay to identify DSIF contained another complex, Negative Elongation Factor (NELF), which acts in concert with DSIF to inhibit elongation. In the absence of DSIF, the NELF complex is unable to bind and repress RNAPII, showing that DSIF is required to inhibit transcription elongation *in vitro* (Yamaguchi et al. 2002). Productive elongation is resumed when P-TEFb (Positive Transcription Elongation Factor) phosphorylates RNAPII and Spt5, thereby dissociating NELF from the TEC.

In addition to inhibiting elongation, DSIF physically associates with multiple components of the transcriptional machinery (TFIIF, TFIIS, CSB, Spt6, FACT, Chd1, Paf complex, Capping enzyme, nuclear exosome) (Sims et al. 2004), suggesting roles in general mRNA processing events related to its stimulatory activity. Furthermore, DSIF has been shown to affect RNAPII processivity (the ability of the elongating RNAPII to travel the entire length of the gene), suggesting the importance of DSIF to the stability of the RNAPII TEC on the chromatin template in yeast (Mason and Struhl 2005). The mechanism of action of Spt5 has been elucidated at the mechanistic level from teasing apart these protein and genetic interaction networks.

In vivo studies in worms, flies and zebrafish have provided considerable evidence for the physiological importance of Spt5 in multi-cellular organisms. Inactivation of Spt5 through RNAi in C. elegans arrests embryonic development prior to gastrulation, but many differentiated cell types are able to form (Shim et al. 2002). In *Drosophila*, a hypomorphic mutation in spt5 results in defects in the expression of segmental patterning genes (Jennings et al. 2004); protein localization on polytene chromosomes and chromatin immunoprecipitation studies reveal broad distribution of Spt5 on the chromatin (Andrulis et al. 2000; Kaplan et al. 2000; Saunders et al. 2003). In zebrafish, hypomorphic mutations in *spt5* (the fog^{m806} allele, the fog^{fh20} allele) display distinct defects in neuronal specification (Guo et al. 2000), and migration (Cooper et al. 2005) respectively, while severe truncation (the fog^{sk8} allele) or deletion (the fog^{s30} allele) of the gene product results in broad deficits in embryonic development (Keegan et al. 2002). These loss-of-function mutants provide the first opportunity to identify target genes of Foggy/Spt5 *in vivo* and in a vertebrate developmental context.

The *sk8* mutants display developmental retardation by 20 hours post fertilization (hpf) followed by further deterioration and death by 48 hpf. The *fog*^{sk8} embryos show several gross abnormalities including pigmentation, tail, ear and cardiac defects. Further analysis of these mutant embryos showed that Foggy/Spt5 is required for normal differentiation of the precardiac mesoderm into myocardial tissue (Keegan et al. 2002). Regardless of these defects, the mutant embryos exhibit normal lens induction, early touch reflex and general body patterning by the late segmentation stage of development. In spite of the initial characterization of the *in vivo* physiological roles of Foggy/Spt5 in different model organisms, many questions remain. A).What genes are regulated by Foggy/Spt5 at the level of transcription elongation? B).What type of elements (*cis* regulatory sequences, gene structure characteristics, function of the genes) may predict Foggy/Spt5's target genes?

Here we report that, through expression profiling studies on ~10,000 protein-coding genes using 24 hours post fertilization (hpf) zebrafish embryos, a surprisingly small fraction (less than 5%) was differentially expressed between fog^{sk8} mutants and their wild-type (WT) siblings. Further classification analyses revealed that these differentially expressed genes are involved in diverse biological pathways ranging from pattern specification to stress response. A permutation based analysis revealed that genes upregulated in *sk8* embryos have significantly shorter gene length compared to the average gene length in zebrafish. Chromatin immunoprecipitation (ChIP) on selected set of genes in WT embryos identified that a large fraction of these genes were occupied by Foggy/Spt5 in a pattern similar to RNAPII occupancy. Furthermore, Foggy and RNAPII occupancies suggest the poised transcriptional status of many upregulated genes, while

many downregulated genes are actively transcribed. Our comparison of RNAPII occupancy between WT and *sk8* embryos identified Growth Arrest and DNA-damage inducible gene (*gadd45b*) to be a novel target for repression by Foggy. Taken together, our results identify a small but functionally diverse set of genes that are regulated by Foggy/Spt5 in vertebrate development and further introduce a novel candidate for regulation by elongation repression.

Materials and Methods:

Zebrafish

Wild-type AB strain and *fog* ^{sk8} heterozygotes were obtained through natural spawning and were maintained and staged according to established procedures (Westerfield, 1993). Embryonic stages at 28.5°C are given in hours post fertilization 9hpf). The *fog* ^{sk8} heterozygote pairs were identified by genotyping and mated to obtain *sk8* mutant embryos. Phenotypic WT siblings (mix of genotypical WT and heterozygotes) were used for most assays while comparing *sk8* embryos. ChIP assays were performed with embryos from WT ABC strain.

Total RNA isolation and qRT-PCR

Equal number of embryos (WT siblings, *sk8* mutants) at 24 hpf were dechorionated by hand or by Pronase K treatment. Appropriate amounts of Trizol were then added to the embryos and stored at -70°C. Standard RNA extraction protocols were followed and total RNA was re-suspended in nuclease-free water. DNase treatment was performed to minimize contamination by genomic DNA during qRT-PCR. To obtain cDNA, total RNA was reverse-transcribed using oligodT and Superscript enzyme. Equal amount of cDNA was used in SybrGreen qPCR mastermix system with appropriate primers. All conditions (primer concentration, and temperature for PCR) were optimized. All plates contained at least two well duplicates for each primer and cDNA condition. Cycle values were averaged between the duplicates and normalized to control primer values. WT fold values were compared to *sk8* mutant fold values and expressed as fold upregulated or downregulated in *sk8* embryos. At least two independent replicates were performed for each primer set.

Microarray Data Analysis

Gene intensity values for the Affymetrix Zebrafish Genome Array CEL files were obtained using the GC-RMA package from Bioconductor (Dudoit S et al 2003; Wu Z et al. 2004). To identify genes with relative changes in steady-state mRNA levels between WT and *sk8* mutants, a fold change was calculated from these log2 intensity values along with a student *ttest p* value. Genes with 2 fold change in expression and $p<0.05$ were clustered using the program HOPACH (hierarchical ordered partitioning and collapsing hybrid), with uncentered correlation distance (Pollard KS et al 2002; van der Laan MJ et al 2003). Associations with Gene Ontology, GenMAPP (Dahlquist et al. 2002) and PANTHER (Mi et al. 2007) pathways were obtained using MAPPFinder 2.0 (Doniger SW et al. 2003), a part of the GenMAPP 2.1 application package and further filtered using the program GO-Elite (*http://www.genmapp.org/go_elite/go_elite.html*).

To determine relative differences in gene length parameters, a custom python script was written that links microarray probe set data to gene structural information from the Ensembl database (PMID: 17148474) obtained through BioMart (Durink S et al, 2005). Probe set annotations to Ensembl genes obtained from BioMart were augmented with those directly from Affymetrix (http://www.affymetrix.com). In this script, fold changes were calculated by comparing the mean length of distinct gene features (genomic, exon and intron lengths/number) or gene expression between upregulated or downregulated Ensembl-linked genes compared to all genes on the microarray. For each parameter, a permutation *p* value was calculated by randomizing the selection of input probe sets 10,000 times and re-calculating and storing the mean fold changes for both up

and down regulated sets to determine the probability that the original fold occurs by chance alone.

Chromatin Immunoprecipitation Assay

WT ABC embryos at 24 hpf were dechorionated by Pronase treatment and washed thoroughly. Embryos were fixed in 1.5% formaldehyde for 20 minutes and then homogenized 25 times in 5X extraction buffer with protease inhibitors on ice. Extracts were then left on the nutator in the cold room for at least 1.5 hours. Sonication was carried out with Branson sonifier 250 on an ice bath for 10 times, 10 seconds each with 20 second breaks. Sonicated samples were then centrifuged at 14k rpm at 4°C for 30 minutes to separate the pellet from the supernatant. The supernatant was immuno-precipitated with appropriate antibodies or controls for 2-4 hours in the cold room. Protein A Sepharose beads were added to the samples and left on overnight. The next day, extensive washing with different buffers were performed before eluting the beads at 65°C for 20 minutes. Proteinase K and RNase was added to the eluates and reverse-crosslinked overnight at 65°C. Phenol/chloroform extraction was performed and followed by ethanol precipitation. The immunoprecipitated DNA was immediately used for qPCRs with appropriate primers. All cycle values were first normalized to input values before comparing control samples to immuno-precipitated samples. These values were represented as fold occupancy over control values. The procedure was performed similarly for *sk8* embryos.

Whole mount staining

Whole-mount *in situ* hybridization was performed as described before. Briefly, embryos were fixed in 4% paraformaldehyde, overnight in the cold room. Then, the

embryos were dechorionated by hand and stored in 100% methanol at -20°C until usage. Embryos were rehydrated and treated mildly with proteinase K before incubation with appropriate digoxigenin-labeled RNA probes. After extensive washing at 60°C, an antibody against digoxigenin is added to the tubes and incubated overnight. NBT/BCIP (Roche) colorimetric reaction is used to visualize the expression patterns of the genes. Plasmid containing *gadd45b* was obtained from ZFIN. Whole-mount immunostaining is performed similarly with appropriate antibodies and fluorescent secondary antibodies. In situ images are taken with Zeiss microscope, while immunostaining images are taken with Zeiss confocal microscope.

Western Blot

Dechorionated embryos were fixed in 1.5% formaldehyde for 20 minutes and homogenized before addition of 2X SDS loading buffer. Equal amounts of the extracts are loaded on to the wells of the gel. After electrophoresis, the proteins on the gel are transferred to a nitrocellulose membrane. After overnight blocking with 5% non-fat milk in PBS-Tween, the membrane is incubated with primary antibodies at appropriate concentrations for 1.5 hours at room temperature. After extensive washing, secondary antibody conjugated to horse-radish peroxidase (HRP) is applied to the membrane for 30 minutes. The ECL kit (Amersham) is used to visualize the protein bands transferred from the blot to film.

Results:

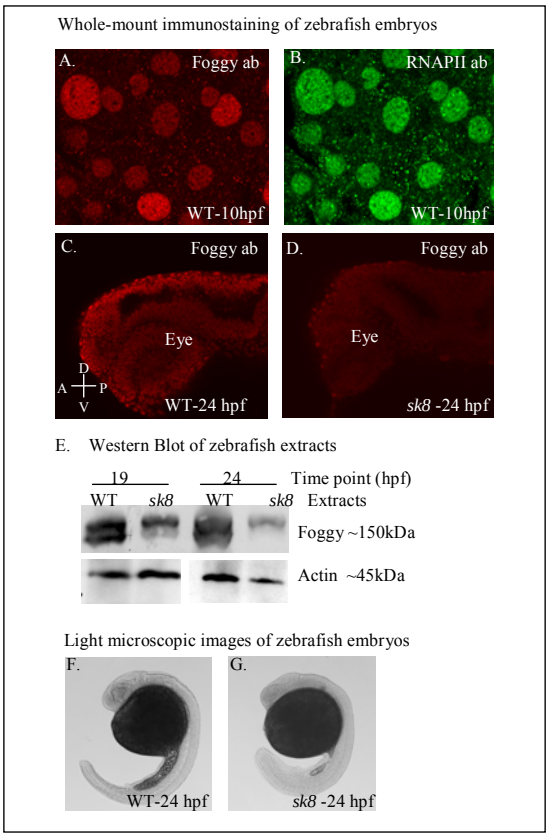


Figure 1: Foggy protein is significantly decreased in *sk8* embryos.
Foggy protein localizes to the nucleus of embryonic cells and is ubiquitously expressed. (A, B). Immunolocalization of foggy (red) and RNAPII (green) in WT embryos at 10 hpf shows colocalization of these two proteins in the nucleus of the embryo. Images are taken with Zeiss confocal microscope at 63X magnification. (C, D). Immunolocalization of foggy (red) in WT and *sk8* embryos at 24 hpf shows overall decrease of Foggy in *sk8* embryos. Confocal images at similar sections of WT and *sk8* embryos are shown at 20X magnification, lateral view. A- anterior, P – posterior, D - dorsal, V- ventral.
Western blot of wild-type (WT) and *sk8* embryo extracts at 19 hpf and 24 hpf using Anti-Foggy antibody shows substantial decrease in maternal Foggy in *sk8* embryos (E). The post-translational modifications (phosphorylation, methylation) of Foggy result in two bands recognized by this antibody. Actin antibody is used as control. Equal numbers of WT and *sk8* embryos were loaded for the western blot, but *sk8* embryos are considerably smaller in size than WT embryos at 24 hpf.
Light microscopic view of WT and *sk8* embryos at 24 hpf (F, G) shows the gross developmental retardation phenotype of the mutant embryos. 5X magnification.

Foggy/Spt5 is significantly reduced in *sk8* mutant embryos at 24 hpf

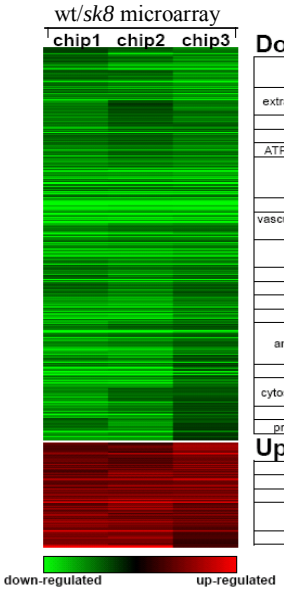
Previous studies (Guo et al. 2000; Keegan et al. 2002) of foggy transcript expression in zebrafish showed a broad distribution throughout the embryo in early stages, but predominantly brain-specific expression by 48 hpf. To gain insights into the sub cellular distribution of Foggy/Spt5 protein, an antibody directed against the C-

terminal end of the zebrafish Foggy/Spt5 protein was generated. Double-labeling with 8WG16, an antibody that recognizes RNAPII, showed predominantly nuclear distribution of Foggy/Spt5 (Fig. 1A, B). The *fog*^{sk8} mutation is predicted to yield a non-functional zygotic gene product, and the maternal transcript is undetectable after 10 hpf (Keegan et al. 2002). However, the status of the maternal protein was unclear. The *sk8* mutant phenotype is first visible around 19 hpf, suggesting the presence of maternal Foggy protein at earlier stages of development. Whole-mount immunostaining and Western blot analysis with Anti-Foggy antibody showed that Foggy protein is significantly reduced in *sk8* embryos by 24 hpf (Fig. 1C, D, E).

Microarray analysis reveals that a small number of genes are differentially expressed in *sk8* embryos at 24 hpf

To identify potential target genes of Foggy/Spt5, we performed microarray analysis by hybridizing zebrafish Affymetrix oligoarrays, which contain 15,617 probesets representing ~10,000 genes, with fluorescently labeled RNAs isolated from *sk8* mutant embryos and their WT siblings at 24 hpf. While 75% of the probe sets display above-background signal-to-noise ratio (as determined by microarray absent-present calls), only 4.5% of the genes are differentially expressed (absolute fold greater than 2, $P < 0.05$) between *sk8* embryos and WT siblings. Among the differentially expressed genes, 97 (~22%) were upregulated (ranging from +2 to +10 fold), and 358 (~78%) were downregulated (ranging from -2 to -300 fold) in *sk8* embryos. Interestingly, many of the genes with altered expression in *fog*^{sk8} embryos have previously been shown to display dynamic spatial and temporal expression patterns (Thisse et al. 2004).

A. Cluster diagram of microarray results with high-ranking GO categories



Down-regulated

GO Name	Number Changed	Number Measured	Z Score	Permute P	GO Type	Corresponding Genes
extracellular matrix structural constituent	5	7	8.03	0	function	*col1a1, col1a2, col1a3, col2a1a, matn4*
tetrapyrrole metabolic process	4	6	6.90	0	process	*alas2, hmbs, hmox1, urod*
Notch signaling pathway	6	15	6.20	0	process	*dla, dlb, dlc, dld, her4.1, notch3*
ATP hydrolysis coupled proton transport	3	5	5.62	0.0015	process	*atp1a1a.4, atp1a2a, atp1a1b*
calcium ion binding	23	167	5.26	0	function	*cdh11, cdh15, cmlc1, dl(a,b,c,d), hsp90b1, lcp1, matn4, myl7, mylz(2,3), notch3, pvalb1, pvalb(2,5,8), tnnc*
oxygen transporter activity	5	16	4.80	0.0005	function	*hbae1, hbae3, hbbe3, mb, vim*
vascular endothelial growth factor receptor activity	3	7	4.58	0.0005	function	*kdr, pdgfra, ttnl*
Neurogenesis	14	113	4.56	0	Panther	*ascl1a, dld, lhx1b, rtn1b, tbr1, tlx3a, tnw, zgc:113283, zgc:64199*
Dr_DNA_replication_Reactome	5	23	4.29	0.001	GenMAPP	*mcm2, mcm3, mcm4, mcm5, orc5l*
porin activity	3	8	4.20	0.003	function	*aqp3, aqp8, zgc:85890*
actin binding	6	27	4.09	0	function	*gsnl1, lcp1, sptb, svil, tmsb, tpma*
Muscle development	5	30	3.50	0.004	Panther	*boc, im:6899248, myl7, pvalb8, ttnl*
anatomical structure morphogenesis	17	159	3.32	0	process	*atoh1a, atp1a1a.2, atp1a2a, boc, ccnd1, cx43, dld, gap43, igfbp1, kdr, lbr, notch3, oep, plasticin, ptf1a, sox9b, tnc*
somitogenesis	5	27	3.21	0.0025	process	*dlc, dld, her4.1, myhz2, oep*
cytoskeleton organization and biogenesis	8	60	2.95	0.002	process	*epb41, prox1, sptb, svil, tmsb, tuba2, tubb5, zgc:112335*
cell migration	6	53	2.10	0.018	process	*boc, lbr, nrarpa, oep, supt5h, tnc*
protein tyrosine phosphatase activity	4	31	2.01	0.041	function	*ptprn2, ptprn, ptpn6, zgc:63553*

Up-regulated

viral envelope	3	4	13.25	0	component	*zgc:114014, zp2.2, zp2.4, zp3.2*
response to cadmium ion	3	8	9.22	0	process	*gadd45b, hsp70, jun*
pattern specification process	5	105	3.27	0.0035	process	*atp1a1a.1, fst, gadd45b, gadd45bl, smo*
sequence-specific DNA binding	13	288	5.14	0	function	*bapx1, fos, foxd5, hox(b6a, b6b, b10a, c5a, 6a, d11a, d13a), jun, nkx2.7, six4.1*
G-protein mediated signaling	3	79	3.28	0.044	Panther	*opn1sw2, smo, tpbgl*

B.

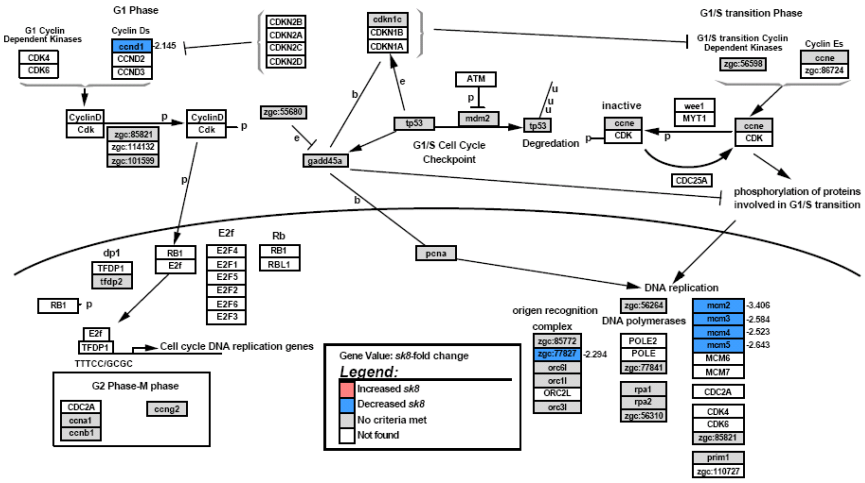


Figure 2: Expression profiling of *sk8* embryos shows distinct categories of differentially expressed genes.

A. Cluster diagram of the microarray data displays considerable similarities across three replicates. 615 probesets with fold change > 2 and ttest p-value < 0.05 are displayed along with Gene Ontology terms and pathways over-represented as determined by a permutation based p-value. Probesets were clustered using the program hopach and the intensity of downregulation (green) and upregulation (red) reflects the magnitude of fold change. Analyzed pathways were obtained from GenMAPP (derived by homology to human) (Salomonis et al, 2007) or Panther Pathways. Representative genes within the GO categories are shown in the last column. Genes which were considered for further analysis in this paper are represented in blue color in the last column.

B. G1 to S Cell Cycle pathway is shown as an example of GenMAPP pathway analysis. The pathway map shows genes that are regulated as a group (*mcm* genes) and genes which are not differentially expressed in *sk8* embryos but are present on the array (grey color).

Genes upregulated in *sk8* embryos				
#	**Probe ID**	**Primers**	**qRTPCR folds**	**Array folds**
1	Dr.592.1.S1_at	Fox D5	10.77	9.83
2	Dr.15991. 1.S1_at	Hypothetical protein	4.40	6.56
3	Dr.1378.2.S1_a_at	Gadd45b	6.07	6.34
4	Dr.12986.2.S1_at	v-fos	8.44	5.68
5	Dr.20198.2.S1_x_at	Hsp70	2.32	5.60
6	Dr.21063.1.A1_at	Bapx	4.62	4.62
7	Dr.15033.1.S1_at	Hypothetical protein	7.33	4.00
8	Dr.13076.1.S1_at	Plekhf1	3.19	3.87
9	Dr.23439.4.S1_at	Zp2.4	6.02	3.74
10	Dr.5725.1.S1_at	HoxB6b	3.84	3.31
11	Dr.25206.1.S1_at	Tpbgl	2.37	3.24
12	Dr.18282.6.A1_at	Khdrbs1	5.94	2.97
13	Dr.8097.1.S1_at	Blue Opsin	7.81	2.72
14	Dr.12334.1.A1_at	Hypothetical protein	4.24	2.35
15	Dr.198.1.S1_at	Fst	2.07	2.25
16	Dr.15833.1.A1_at	Rpb1	3.71	2.17
17	Dr.24766.1.S1_at	Smo	2.12	2.02
18	Dr.14282.1.S1_at	ATF3	1.89	2.00
19	Dr.24669.1.S1_at	TFIIA	3.61	1.82
Genes downregulated in *sk8* embryos				
1	Dr.14668.1.S1_at	Gch	899.50	320.56
2	Dr.20928.1.S1_at	Parvalbumin 1d	21.47	22.35
3	Dr.23350.1.S1_at	Parvalbumin 8	12.23	17.62
4	DrAffx.1.52.S1_at	A2BP1-like	7.83	10.69
5	Dr.12107.1.A1_at	Ndrg1	8.47	6.04
6	Dr.11483.1.S1_at	Ldb3-like	3.44	3.64
7	Dr.10719.1.S1_at	Atp1a1a.4	62.85	3.37
8	Dr.1831.1.S1_at	Lfng	3.68	3.00
9	Dr.20815.1.S1_at	Tpma	3.41	2.86
Control genes				
1	Dr.25213.1.S1_at	Beta Actin 1	-1.80	1.26
2	Dr.1347.1.S1_at	Ribosomal protein L4	1.00	1.00
3	Dr.2675.1.A1_at	FKBP5	1.00	1.04
4	Dr.29.1.S1_at	CCNE	1.87	1.03

Table 1: Most differentially expressed genes from the microarray are validated independently with qRT-PCR. Total RNA was extracted from equal number of WT and *sk8* embryos at 24 hpf and reverse transcribed with oligodT. Primer pairs were designed towards the 3' end of the genes using Ensembl sequence information. qPCR with Sybr Green was performed using appropriate primers. Averages from at least two independent repetitions for each primer pair is represented in the table.

Nineteen upregulated genes, nine downregulated genes and four unchanged genes were further verified by quantitative RT-PCR in WT and *sk8* embryos at 24 hpf. Differential expression was confirmed for all upregulated and downregulated genes examined; one of the unchanged genes (bactin 1) showed a minor decrease (-1.8) in *sk8* embryos (Table 1).

To identify biological processes that might be regulated as a group by Foggy/Spt5, we employed GenMAPP (Dahlquist et al. 2002) and associated tools to analyze pathway and gene ontology over-representation in our dataset (Fig. 2A). This analysis showed that genes that are constituents of the extracellular matrix (e.g., *collagen*) and sodium/potassium ion transporters (e.g., *atp1a1a*) were downregulated, while genes that are induced by environmental stress (e.g., *gadd45b, hsp70*), and transcription factors (e.g., *foxd5, fos*) were upregulated in *sk8* mutants. In addition, GenMAPP analysis revealed that the G1 to S cell cycle pathway was significantly affected in *sk8* embryos (Fig. 2B).

A number of genes known to be regulated at the level of elongation (Uptain et al. 1997) were also differentially expressed in *sk8* embryos (*hsp70, fos, globin, tubulin, opsin*). As the microarray was performed during early developmental stages, we paid particular attention in our analysis to major developmental pathways. Some members of the Notch signaling pathway (6/15) were downregulated in sk8 embryos while genes in fibroblast growth factor (FGFs), bone morphogenetic protein (BMPs) and Wingless (Wnts) pathways were not differentially expressed, suggesting that these pathways are largely unaffected by lack of Foggy/Spt5. Furthermore, RNA processing genes and protein biosynthesis genes (ribosome constituents) were under-represented in our dataset,

suggesting that downstream general mRNA processing and translation events were not affected in *sk8* mutants. Thus, expression profiling of WT and *sk8* embryos at 24 hpf reveals that Foggy/Spt5 is essential for the regulated expression of a small number of genes during vertebrate development.

Genes with shorter lengths are specifically affected by lack of Foggy/Spt5

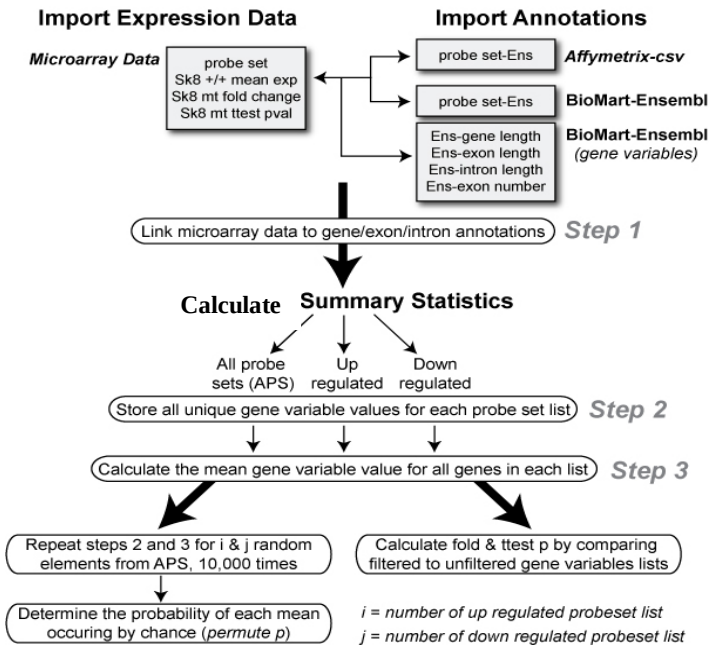


Figure 3: Flow chart diagram of the permutation based analysis
To assess the extent and overall likelihood of a difference between the length of Foggy/Spt5 regulated genomic structure to other genes in the array, we linked microarray probe sets to gene structure data from the Ensembl database, computed the mean of different gene features (e.g., genomic length) for differentially expressed genes and generated fold and *P* values.

Previous studies in yeast have suggested that for intron-containing genes, Foggy/Spt5 may couple regulation of transcription elongation with splicing (Xiao et al. 2005). To determine if differentially expressed genes in *sk8* embryos have specific gene structure features, we performed bioinformatics analysis on these gene sets. To assess the extent and overall likelihood of a difference between the length of Foggy/Spt5-regulated genomic structure to other genes in the array, we linked microarray probe sets to gene structure data from the Ensembl database, computed the mean of different gene features (e.g., genomic length) for differentially expressed genes and generated fold and P values using a permutation based analysis (Fig. 3).

Categories	**Total**	**Upregulated in *sk8* embryos**			**Downregulated in *sk8* embryos**		
Groups	Mean	Mean	Fold	PermuteP	Mean	Fold	PermuteP
Gene length	18233	7618	**-2.39**	0	18468	1.01	0.8576
Exon length	182	228	**1.26**	0.0158	192	1.06	0.1648
Exon number	10	7	-1.45	0.8956	11	1.12	0.7372
Intron length	2165	1277	**-1.69**	0.0002	1850	-1.17	0.0516
Average expression	293	53	-5.55	0.0852	579	1.98	0
Genes	9955		97			358	
Probesets	15617		129			485	

Table 2: Summary of the permutation based analysis of the gene structure requirements of differentially expressed genes in *sk8* embryos. Upregulated genes have significantly shorter gene length, shorter intron length and longer exon length than other genes on the array. Downregulated genes do not show significantly different gene structure requirements compared to other genes.

Interestingly, genes upregulated in *sk8* embryos were found to have on average ~2.4 fold decrease in overall gene length and ~1.7 fold decrease in intron length compared to all gene-linked probe sets on the array ($p<0.05$) (Table 2). The extent of this difference was not observed over 10,000 random permutations of the data and was

observed only twice when comparing intron length (probabilities of 0 and 0.0002 respectively), emphasizing the rare likelihood of such events. Although this analysis also shows a significant increase in exon length for upregulated genes ($p<0.01$), the difference is relatively small (1.26 fold). In contrast to these observations for upregulated genes, none of the gene structure data differed between downregulated genes and all other genes on the array. Our permutation analysis of differentially expressed genes suggests that genes that are smaller in length and have shorter introns might be repressed by Foggy/Spt5.

Distinct patterns of RNAPII occupancy is detected on candidate genes in WT embryos at 24 hpf

As microarray experiments measure the steady-state levels of mRNA and Foggy/Spt5 is involved in many aspects of mRNA stability by virtue of its association with many components of the transcriptional machinery (Sims et al. 2004), many of the genes differentially expressed in *sk8* mutants might represent mRNA processing targets of Foggy/Spt5. In order to better identify genes that are regulated at the elongation level at 24 hpf, we performed chromatin immunoprecipitation (ChIP) assays using antibodies against RNAPII and Foggy. Candidate genes were picked based on fold change, reliable genomic structure (exon/intron) information, and function that might elucidate the *sk8* mutant phenotype. The fold change and genomic structure information for the ChIP candidate genes is presented in Table 3.

No.	Name	GO /MAPP	Array Fold	Genomic length (kb)	Transcript length (kb)	No. of Exons
Upregulated genes						
1	FoxD5	Transcription Factor	9.83	4.1	1.1	2
2	Gadd45b	Stress response	6.34	1.6	1.2	4
3	Fos	Transcription Factor	5.68	2.1	1.7	4
4	Hsp70	Stress response	5.60	3.4	2.3	2
5	Bapx	Transcription Factor	4.62	0.5	0.3	2
6	Zp2.4	Viral envelope	3.74	1.8	1.3	8
7	Tpbgl	Receptor	3.24	2.6	1.7	2
8	Opsin	Signal transduction	2.72	2.6	1.3	5
9	Fst	Pattern specification	2.25	5.5	3.4	5
10	Smo	Pattern specification	2.02	5.6	3.0	12
11	ATF3	Transcription Factor	2.00	6.1	2.6	4

Downregulated Genes						
No.	Name	GO /MAPP	Array Fold	Genomic length (kb)	Transcript length (kb)	No. of Exons
1	Pvalb8	Calcium ion binding	17.62	174.7	0.8	5
2	A2BP1l	Nucleotide binding	10.69	18.5	1.1	12
3	Ndrg1	Cell differentiation	6.04	19.9	2.3	16
4	Ldb3l	Protein binding	3.64	39.3	1.6	12
5	Atp1a1a.4	Sodium\potassium ATPase	3.37	7.5	3.2	22
6	Lfng	Signaling	3.00	9.4	2.9	8
7	Tpma	Actin binding	2.86	4.5	1.3	10

Control Genes						
1	Bactin1	Cytoskeleton	1.26	3.6	1.7	6
2	CCNE	Cell cycle	1.00	9.8	2.0	12
3	FKBP5	Protein folding	1.00	9.9	2.4	11

Table 3: GO terms suggesting function, microarray fold change, and gene structure information on genes selected for chromatin immunoprecipitation assay with Anti-RNAPII and Anti-Foggy antibodies in WT embryos at 24 hpf. Most of the upregulated genes are shorter than the average gene length while downregulated genes cover a wider range of gene lengths.

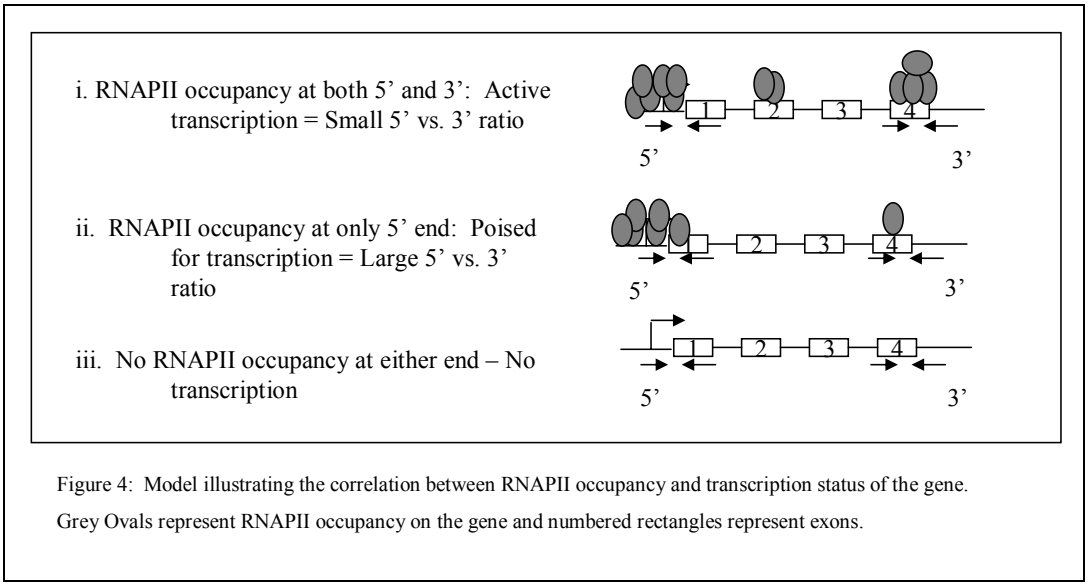


Figure 4: Model illustrating the correlation between RNAPII occupancy and transcription status of the gene. Grey Ovals represent RNAPII occupancy on the gene and numbered rectangles represent exons.

We decided to use 8WG16 antibody to detect RNAPII because this antibody was successfully used to recognize RNAPII occupancy on the chromatin at both the 5' and 3' ends of genes in other species (Kim et al. 2004). The prevailing model correlating RNAPII occupancy to transcriptional status on the gene is schematically depicted in Figure 4. RNAPII occupancy at both 5' and 3' ends of the gene is associated with productive transcription, while lack of detectable RNAPII on the gene is associated with no transcription at the time point tested. The intermediate status of considerable RNAPII occupancy at 5' end and minimal occupancy at 3' end is associated with a gene poised for transcription.

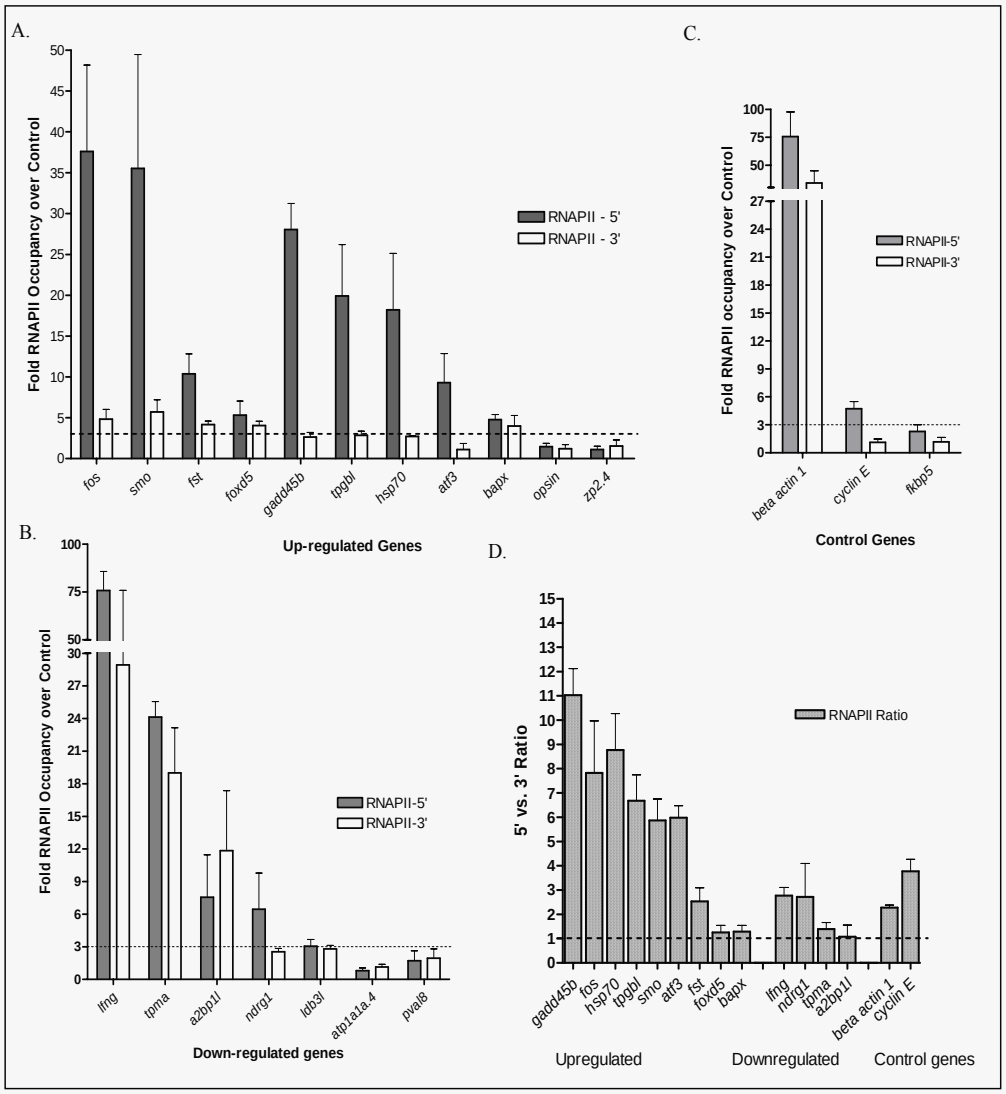


Figure 5: Distinct patterns of RNAPII occupancy indicating transcriptional status detected on selected set of genes in WT embryos at 24 hpf. RNAPII occupancy on (A) upregulated genes, (B) downregulated genes, (C) and control genes; (D) RNAPII occupancy at 5' and 3' end represented as ratio. ChIP was performed using 8WG16 antibody that recognizes RNAPII on the chromatin. QPCR with appropriate primers was performed in duplicates and normalized to input DNA values. Y axis values in the graphs represent fold RNAPII occupancy over fold mouse control serum occupancy. Arbitrary cutoff for detectable RNAPII fold occupancy over control serum is set stringently at 3. The results shown are the average of at least three independent replicates. Error bars are SEM values.

Most of the upregulated genes (9/11) showed RNAPII occupancy at 5' end of the gene in WT (Fig. 5A). Of the nine genes, four genes (*fos, smo, fst, foxd5*) show RNAPII occupancy at the 3' end as well. Based on the model, these genes are engaged in productive transcription whereas the other five (*gadd45b, tpgbl, hsp70, atf3, bapx*) are in a poised state. Genes without detectable RNAPII on the gene (*opsin, zp2.4*) are nottranscribed at this time point.

Three (*lfng, tpma, a2pb1l*) of seven downregulated genes show detectable occupancy at 5' and 3' end, suggesting productive transcription of these genes. The other four genes (*ndrg1, lbd3l, atp1a1a, pvalb8*) show minimal occupancy at either end, suggesting lack of transcription at this time point (Fig. 5B).

Genes that do not show significant changes in mRNA level between WT and sk8 embryos (*bactin1, cyclin E, fkbp5*) were used as controls; these three genes also show the three patterns of RNAPII occupancy (Fig. 5C). Beta actin 1 shows considerable RNAPII occupancy at 5' and 3' end (active transcription), cyclin E shows occupancy at 5' end (poised for transcription), and fkbp5 does not show detectable RNAPII occupancy at either end (no transcription).

When RNAPII occupancies were analyzed as ratio within the gene, all the downregulated genes (4/4) have small ratios, indicating active transcription (Fig. 5D). Given that these actively transcribing genes are downregulated in *sk8* embryos, it is likely that they require the stimulatory activity of Foggy/Spt5. In contrast, seven of the nine upregulated genes show high ratios, suggesting that these genes might be repressed by Foggy/Spt5.

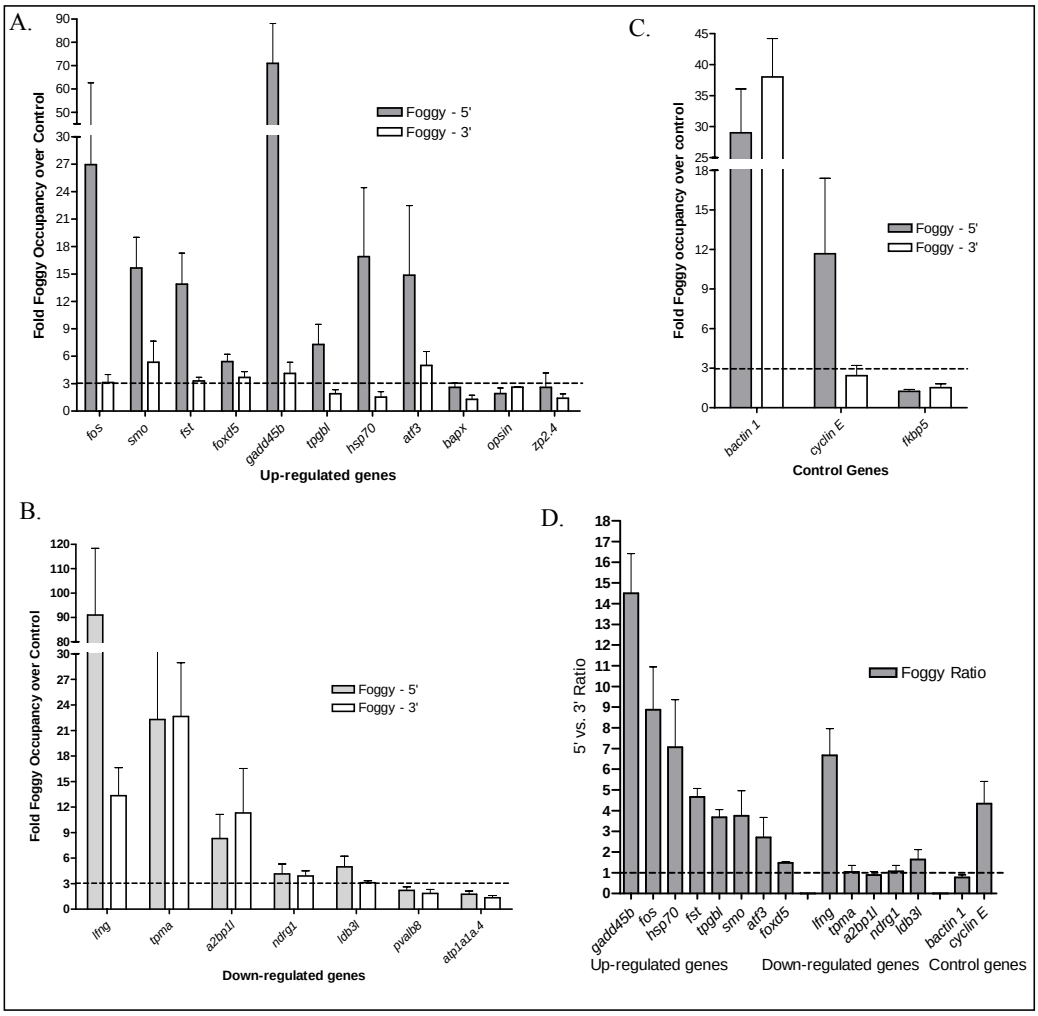


Figure 6: Foggy occupancy mirrors RNAPII occupancy on majority of selected genes in WT embryos at 24 hpf. Foggy occupancy on (A) upregulated genes, (B) downregulated genes, (C) and control genes; (D) Foggy occupancy at 5' and 3' end represented as ratio. ChIP was performed using Anti-Foggy antibody and pre-immune serum in procedure similar to that used for RNAPII ChIP. Y axis values in the graphs represent fold Foggy occupancy over fold pre-immune serum occupancy levels. Arbitrary cutoff for detectable Foggy fold occupancy over control serum is set stringently at 3. The results shown are the average of at least three independent replicates. Error bars are SEM values. Note that the scales are different for each of the graphs.

Foggy/Spt5 occupancy mirrors RNAPII occupancy on most candidate genes in WT embryos at 24 hpf

Our analysis of selected candidate genes using RNAPII antibody identified distinct patterns of RNAPII occupancy and consequently, the transcription status of the gene at 24 hpf. Next, in order to identify genes that are regulated at the elongation level by Foggy/Spt5 at 24 hpf, we performed chromatin immunoprecipitation (ChIP) assays using anti-Foggy antibody. In WT embryos, most of the upregulated genes (8/11) show Foggy occupancy at 5' end and minimal occupancy at 3' end of the genes. Three of the upregulated genes (*bapx, opsin, zp2.4*) do not show detectable Foggy occupancy at this particular time point (Fig. 6A). *Bapx*, which showed RNAPII occupancy, does not show above-threshold Foggy occupancy suggesting that Foggy/Spt5 is not essential for transcribing this gene. Most of the upregulated genes that are poised for transcription (as assessed by RNAPII occupancy) do display similar Foggy occupancy at 5' end of gene, suggesting that transcription of these genes might be repressed by Foggy protein.

Three out of seven downregulated genes (*lfng, tpma, a2bp1l*) show Foggy occupancy at both 5' and 3' ends, similar to RNAPII occupancy on these genes (Fig. 6B). It is likely that Foggy/Spt5 plays a stimulatory role in transcribing these genes, resulting in their downregulation in *sk8* embryos. Two of the downregulated genes (*atp1a1a.4, pvalb8*) show minimal Foggy occupancy at both ends, consistent with these genes not being transcribed at this time point. Foggy occupancy on control genes (*beta actin 1, cyclin E, fkbp5*) is similar to RNAPII occupancy (Fig. 6C). Given that these genes are not differentially expressed in *sk8* embryos, it is likely that these genes are not uniquely regulated by Foggy/Spt5. When occupancy values were computed as ratios, most of the

upregulated genes have high ratios and most of the downregulated genes have low ratios (Fig. 6D), similar to RNAPII ratios (Fig. 5D).

Taken together, these results identify *gadd45b, fos, hsp70, smo, atf3, fst*, and *tpgbl* as promising candidate genes for direct repression by Foggy/Spt5 at the elongation step, and *lfng, tpma*, and *a2bp1l* as promising candidate genes for direction stimulation by Foggy/Spt5.

Identification of *gadd45b* as a novel elongation target of Foggy/Spt5

We hypothesized that if a gene loses its elongation repression in *sk8* embryos, there would be an increase of RNAPII occupancy at the 3' end of the gene, reflecting an increase in mRNA transcription. In order to distinguish between transcription initiation and elongation effects, our hypothesis also predicts that for such genes, there would be no significant increase of RNAPII occupancy at 5' end in *sk8* embryos. Due to the limited availability of fog^{sk8} embryos, we focused on four genes (*hsp70, fos, gadd45b, fst*) which were upregulated in fog^{sk8} embryos, and exhibited strong occupancy of both Foggy/Spt5 and RNAPII at the 5' end. Both *hsp70* and *fos* are known to be regulated by transcription elongation in cultured cells and/or in *Drosophila* (Collart et al. 1991; Gilmour and Lis 1986), whereas *gadd45b* and *fst* are novel candidates for regulation at the elongation step. The patterns of RNAPII occupancy in 24 hpf fog^{sk8} embryos are shown in Fig. 7A.

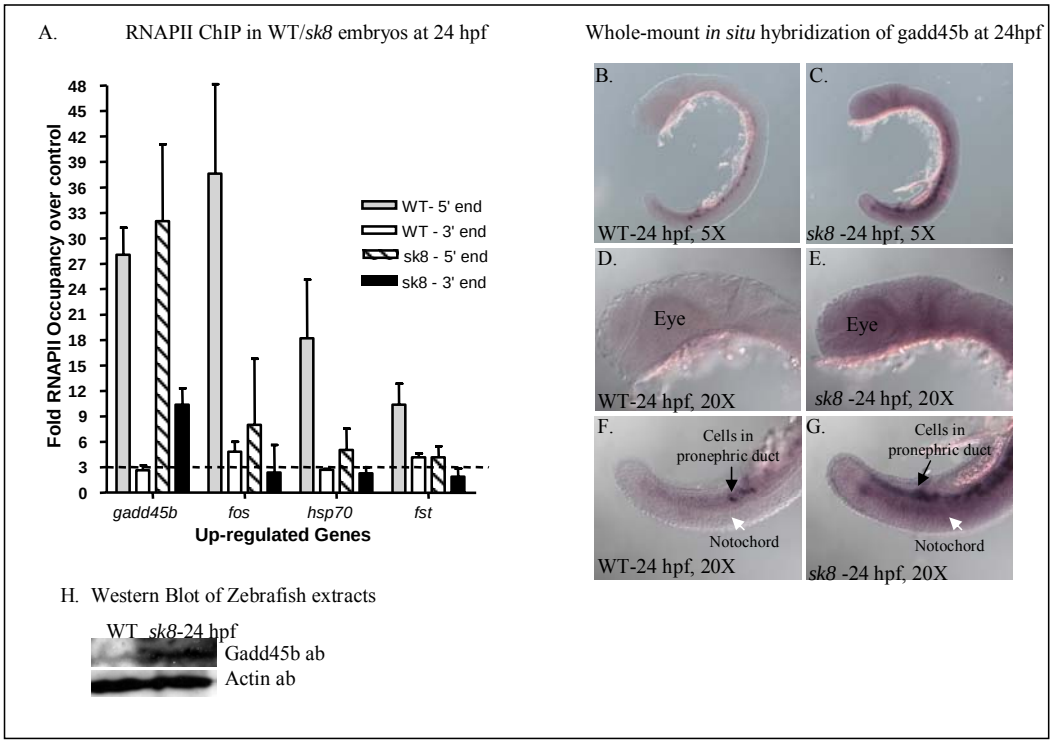


Figure 7: *gadd45b* is repressed by Foggy/Spt5 at the elongation level.
(A). ChIP with RNAPII antibody in WT and *sk8* embryos on *gadd45b*, *fos, hsp70* and *fst* genes show that in *sk8* embryos, RNAPII occupancy is significantly increased at 3' end of *gadd45b* without increase at 5' end. Other genes tested showed reduction of RNAPII occupancy at 5' ends in *sk8* embryos compared to WT occupancy levels. (B-G) Representative images of whole-mount in situ hybridization of *gadd45b* antisense probes in WT and *sk8* embryos at 24 hpf show widespread expression of *gadd45b* in *sk8* embryos in contrast to the limited expression in WT embryos. (H) Western blot analysis of WT and *sk8* embryo extracts with commercially available anti-Gadd45b antibody and anti-actin antibody shows significant increase of Gadd45b protein in *sk8* embryos at 24 hpf. Equal number of embryos were loaded in each lane.

Interestingly, we observed the expected pattern of significantly increased RNAPII occupancy at only the 3' end of *gadd45b* and not with *hsp70, fos* and *fst* genes in *sk8* embryos. Furthermore, we observed a significant decrease of RNAPII occupancy at the 5' end of *hsp70, fos* and *fst* genes when compared to the WT RNAPII occupancy levels. While reasons behind the lack of predicted RNAPII occupancy on *hsp70, fos* and *fst*

needs to be investigated further, the RNAPII occupancy observed for *gadd45b* is consistent with it being a direct target gene repressed by Foggy/Spt5.

Gadd45b mRNA is mis-expressed spatially in *fog* ^{sk8} embryos at 24 hpf

In WT embryos, RNAPII occupancy is asymmetrically concentrated at 5' end of the *gadd45b* gene with minimal 3' occupancy, correlating with minimal expression of gadd45b mRNA at this stage. We hypothesized that the de-repression of *gadd45b* in *sk8* embryos would result in its mRNA being expressed in parts of the embryo that normally do not express *gadd45b*. In order to test this hypothesis, we performed whole mount *in situ* hybridization with *gadd45b* probes in a clutch of embryos from adult heterozygote pairs. The genotype of this the clutch of embryos consists of 25% *sk8* mutants, 50% heterozygotes and 25% WT embryos. At the phenotypic level, heterozygotes and WT embryos are indistinguishable. The spatial expression pattern of *gadd45b* in this clutch is normal in all embryos (regardless of genotype) until 21 hpf. At 24 hpf, phenotypically normal embryos express gadd45b mRNA in a restricted fashion in some cells of the pronepric duct. In *fog*^{sk8} embryos at 24 hpf, we see strong, broad expression in the trunk (presomitic mesoderm, notochord, pronephric duct) and in the anterior parts (layers of the retina, presumptive mid-hind brain boundary, and anterior part of the spinal cord) of the embryos (Fig. 7B-G).

To determine whether increased production of *gadd45b* mRNA in *fog*^{sk8} embryos leads to increased level of the protein product, Western blot analysis was carried out using a commercially available anti-Gadd45b antibody (Fig. 7H). Indeed, a strong increase in Gadd45b protein was detected in 24 hpf *fog*^{sk8} embryo extracts as compared to

WT, suggesting that mis-expression of gadd45b mRNA at the elongation level leads to increased Gadd45b protein levels. Taken together, our data suggests that the loss of repressive activity of Foggy/Spt5, due to decreasing levels of maternal Foggy protein leads to spatial and temporal mis-expression of gadd45b mRNA and protein in *sk8* mutant embryos.

Gadd45b induction in *sk8* embryos is specific to Foggy/Spt5 reduction

Gadd45b is induced in response to various stressful stimuli. In order to determine whether *gadd45b* is induced in other mutant embryos, we analyzed zebrafish microarray data from another mutant (*spt6* mutant) which has severe developmental defects similar to *sk8* embryos at 24 hpf (Len Zon's lab, unpublished data). Spt6, a chromatin remodeling and transcription elongation factor, is associated biochemically with Spt5/Foggy. In fact, the zebrafish mutants *pandora* (*spt6* null allele) and *sk8* (*spt5* zygotic null) were identified in the same screen and showed gross morphological similarities in phenotype (Keegan et al. 2002). The *spt6* mutant allele from the Zon lab has comparable, severe developmental phenotype to *sk8* embryos. Our comparison of the microarrays between the two mutants shows that 67% of the differentially expressed genes in *sk8* embryos are similarly mis-expressed in the *spt6* mutant as well (Table 4). Of the genes that we performed RNAPII ChIP in *sk8* embryos, (*gadd45b, hsp70, fos, fst*), only *fos* is similarly increased while *hsp70* and *fst* are decreased in the *spt6* mutant embryos. Interestingly, *gadd45b* is not differentially expressed in *spt6* mutant embryos, suggesting that *gadd45b* is not a target of *spt6*. This comparison of the two arrays

suggests the extent of the similarities and differences between the targets of these two transcription elongation factors.

Genes upregulated in *sk8* embryos				
#	**Probe ID**	**Primers**	***sk8* Array folds**	*spt6* mutant array folds
1	Dr.8097.1.S1_at	Blue Opsin	2.72	10.82
2	Dr.23439.4.S1_at	Zp2.4	3.74	5.01
3	Dr.592.1.S1_at	Fox D5	9.83	4.86
4	Dr.21063.1.A1_at	Bapx	4.62	4.8
5	Dr.12986.2.S1_at	v-fos	5.68	4.17
6	Dr.15033.1.S1_at	Hypothetical protein	4.00	4.1
7	Dr.13076.1.S1_at	Plekhf1	3.87	2.5
8	Dr.5725.1.S1_at	HoxB6b	3.31	2.34
9	Dr.24766.1.S1_at	Smo	2.02	2.27
10	Dr.25206.1.S1_at	Tpbgl	3.24	2.01
11	Dr.18282.6.A1_at	Khdrbs1	2.97	1.74
12	Dr.15991. 1.S1_at	Hypothetical protein	6.56	1.21
13	Dr.1378.2.S1_a_at	Gadd45b	6.34	1.08
14	Dr.12334.1.A1_at	Hypothetical protein	2.35	1.02
15	Dr.14282.1.S1_at	ATF3	2.00	-1.14
16	Dr.24669.1.S1_at	TFIIA	1.82	-1.24
17	Dr.15833.1.A1_at	Rpb1	2.17	-1.49
18	Dr.198.1.S1_at	Fst	2.25	-1.85
19	Dr.20198.2.S1_x_at	Hsp70	5.60	-2.6

Genes downregulated in *sk8* embryos				
1	Dr.14668.1.S1_at	Gch	320.56	72.56
2	DrAffx.1.52.S1_at	A2BP1-like	10.69	18.53
3	Dr.20928.1.S1_at	Parvalbumin 1d	22.35	16.39
4	Dr.10719.1.S1_at	Atp1a1a.4	3.37	9.96
5	Dr.20815.1.S1_at	Tpma	2.86	8.98
6	Dr.11483.1.S1_at	Ldb3-like	3.64	6.54
7	Dr.23350.1.S1_at	Parvalbumin 8	17.62	6.44
8	Dr.12107.1.A1_at	Ndrg1	6.04	6.21
9	Dr.1831.1.S1_at	Lfng	3.00	2.47

Control genes				
1	Dr.25213.1.S1_at	Beta Actin 1	-1.26	-1.19
2	Dr.1347.1.S1_at	Ribosomal protein L4	1.01	1.01
3	Dr.2675.1.A1_at	FKBP5	1.04	1
4	Dr.29.1.S1_at	CCNE	1.03	1.14

Table 4: Comparison of microarray fold data of select set of genes between *sk8* and *spt6* mutants at 24 hpf. All of the selected downregulated genes show similar fold decreases in both arrays, while 67% of the upregulated genes are similar in fold increases between the mutants.

Foggy/Spt5 might be required to maintain RNAPII conformation on the chromatin

In order to investigate the lack of predicted RNAPII occupancy in *sk8* embryos on majority of the genes tested thus far, we performed whole-mount immunostaining of WT and *sk8* embryos using 8WG16 antibody. 8WG16 antibody recognizes different forms of RNAPII (regardless of the phosphorylation status) between 210-250kDa, (Bregman et al. 1994; Ni et al. 2004). Surprisingly, we detected a significant reduction in RNAPII immunoreactivity by 24 hpf in *sk8* embryos (Fig. 8 A-F). In contrast, the reduction in RNAPII protein levels as detected by Western Blot was not significant (Fig. 8M). To further investigate this issue, we performed immunostaining with another RNAPII antibody, CTD4H8, and found that the RNAPII immunoreactivity in *sk8* embryos was not affected (Fig. 8 G-L). CTD4H8 antibody preferentially recognizes the Ser-5 phosphorylated form of RNAPII (Covance Research Products). This lack of reduction in immunoreactivity by CTD4H8 antibody and Western Blot with 8WG16 antibody suggests that total RNAPII protein levels are not reduced in *sk8* embryos. Preliminary experiments were carried out using another RNAPII antibody (H5), which recognizes the elongating RNAPII (Ser 2-phosphorylated). So far, the results are inconclusive due to technical problems associated with different batches of the antibody. Also, this antibody requires an IgM secondary antibody, for which the conditions have not been optimized thoroughly.

One possible explanation for these discrepancies in the levels of RNAPII in this case is that the two RNAPII antibodies recognize different conformations of RNAPII, and the 8WG16-recognizable RNAPII conformation is specifically reduced in *sk8* embryos. Given that SDS-PAGE gels resolve denatured forms of the protein, the

conformation-specific reduction of RNAPII identified by 8WG16 antibody in the whole-mount immunostaining would not be expected in the Western blot. Nevertheless, the gross reduction of 8WG16-recognizable RNAPII in *sk8* embryos at 24 hpf as detected by whole-mount immunostaining methods, explains the loss of RNAPII occupancy on majority of genes tested by ChIP in *sk8* embryos at 24 hpf. Taken together, these results suggest that Foggy/Spt5 might be required to stabilize particular conformations of RNAPII and consequently, RNAPII processivity on the chromatin.

Discussion:

Foggy/Spt5 is a transcription elongation factor with dual properties of repression and activation *in vitro* (Wada et al. 1998) and *in vivo* (Compagnone-Post and Osley 1996; Hartzog et al. 1998; Winston and Sudarsanam 1998) (Andrulis et al. 2000; Kaplan et al. 2000; Wu et al. 2003). Spt5 also binds multiple mRNA processing factors, suggesting roles in mRNA stability and transport. However, the prevalence and identity of target genes are poorly defined in the vertebrate system. In this study, we carried out expression profiling on ~10,000 genes using a zygotic null zebrafish mutant, *fog*^{sk8}. Maternally deposited *foggy/spt5* gene products sustain *fog*^{sk8} embryos until the late somitogenesis stage of development. Our results show that a surprisingly small fraction of genes (<5%) are differentially expressed between WT and *fog*^{sk8} embryos at 24 hpf. *In vivo* ChIP analyses with anti-RNAPII and anti-Foggy antibodies on a select set of genes established the transcriptional status of these genes and identified possible direct targets of Foggy/Spt5 at the level of transcription elongation at 24 hpf. One such target is *gadd45b*, an anti-apoptotic protein, induced by environmental and cellular stress.

The prevalence and identity of Foggy/Spt5 target genes in vertebrate development

Previous immunofluorescence analyses in Drosophila showed that RNAPII and Spt5 were localized to majority of sites on polytene chromosomes (Andrulis et al. 2000; Kaplan et al. 2000). These studies implied that most, if not all, genes may be dependent on Foggy/Spt5 for productive transcription. On the other hand, the ubiquitous localization and functional redundancy of transcription elongation factors implies that most genes might not be critically dependent on one factor for proper elongation. Our

survey of ~10,000 genes at one developmental stage (24 hpf) in zebrafish embryos reveals that less than 5% of expressed genes are altered in *fog*^{sk8} embryos, suggesting that only these genes require Foggy/Spt5 for their regulation at this particular time point. If the array fold cut-offs were relaxed to an absolute value of 1.5, then the number of upregulated probe sets increase to 693 (129 probe sets at 2-fold cut-off) and downregulated probe sets increase to 1006 (485 probe sets at 2-fold cut-off). Preliminary attempts to validate candidate genes around the 1.5 fold cut-off using qRT-PCR were unsuccessful due to unreliable fold changes between replicates.

Our analyses do not exclude the possibility of larger number of Foggy/Spt5 target genes during other time points of development and/or with a more encompassing microarray chip. As to the identities of Foggy/Spt5 target genes, a permutation based analysis revealed that genes upregulated in *sk8* embryos had significantly shorter gene length compared to the average gene length in zebrafish, suggesting an additional repressive level of regulation for transcription of shorter genes. We did not observe genes downregulated in *sk8* embryos to have significantly longer gene lengths, suggesting that the stimulatory activity of Foggy/Spt5 might be compensated for by other transcription elongation factors. We also identified a diverse set of genes, ranging from pattern specification factors to transporters to be affected by the loss of Foggy/Spt5.

Dynamics of regulated transcription elongation *in vivo*

Our ChIP assays with RNAPII and Foggy/Spt5 antibodies enabled us to classify the differentially expressed genes based on their transcriptional status. In agreement with previous studies, inducible genes, which respond to stimuli quickly in a controlled

manner, were poised for transcription in our assay. These genes were also upregulated in *sk8* embryos, suggesting that these genes are repressed by Foggy/Spt5. Well-known transcription elongation targets, *hsp70* and *fos*, showed predicted Foggy/Spt5 and RNAPII occupancies in WT embryos, though we detected unexpected decrease in RNAPII occupancy of *hsp70* and *fos* genes in fog^{sk8} embryos. One possible explanation for this unexpected finding is that these genes might be more sensitive to reducing amounts of maternal Foggy/Spt5 before 24 hpf, and thus were transcribed at earlier time points. If this was the case, we would expect RNAPII occupancy on *hsp70* and *fos* genes in *sk8* embryos at 24 hpf to resemble WT occupancy patterns (increased RNAPII at 5' end), while we see the contrary pattern (decreased RNAPII at 5' end).

Another explanation for decreased RNAPII could be that the complex is in an arrested state. When RNAPII cannot recover from transcriptional arrest, it is ubiquitylated and degraded. It is interesting to note that the Ser-5 phosphorylated form of RNAPII (close to promoter) is refractory to such ubiquitylation, while Ser-2 phosphorylated form can be ubiquitylated, suggesting that elongating RNAPII is subjected ubiquitylation and possible degradation (Somesh BP et al. 2005). Though the mechanism is unclear, it is also possible that the ubiquitylation serves to provide another level of modification on the CTD rather than serving as a marker for protein degradation. It is also unknown if the RNAPII conformation and the ability of the antibodies used to detect RNAPII is affected by such modification.

Spt4-5 complex is important for stabilizing RNAPII on the chromatin template under normal and DNA-damage inducing conditions in yeast (Mason and Struhl 2005), (Jansen et al. 2002). It is likely that Foggy/Spt5 is important for stabilizing RNAPII on

the chromatin of vertebrates as well, suggesting that RNAPII might be considered a target of Foggy/Spt5. Analysis using antibodies recognizing different conformations of RNAPII would further elucidate the mechanism by which Foggy/Spt5 stabilizes particular conformations of RNAPII. Our ChIP analyses in *fog*^{sk8} embryos have provided further evidence for *gadd45b* as a novel target of Foggy/Spt5's elongation repressive activity. The significantly upregulated Gadd45b protein levels also substantiate the importance of Foggy/Spt5's repressive activity *in vivo*.

Regulation and function of *gadd45b*

A recent study in flies (Peretz et al. 2007), has identified that while overexpression of D-GADD45 in somatic follicle cells caused apoptosis, overexpression in the germline affected the dorsal-ventral polarity of the eggshell and disrupted the localization of anterior-posterior polarity determinants, suggesting roles in early development. In zebrafish, overexpression of *gadd45b* by mRNA injection at the 1-cell stage leads to somite segmentation defects (down-regulation of myoD) (Kawahara et al. 2005). As *gadd45b* is not overexpressed in *sk8* embryos until 24 hpf, it is interesting to note that *sk8* embryos do not show obvious segmentation defects and no reduction in *myoD* transcript levels, suggesting that temporal and spatial context greatly influence the phenotypic outcome of *gadd45b* overexpression during development. It is possible that *gadd45b* upregulation creates a cellular stasis condition in *sk8* embryos, at least for a limited amount of time, in accordance with its role in cell cycle control. Interestingly, our GenMAPP analysis suggested that the cell cycle pathway was indeed affected in *sk8*

embryos. Further in-depth studies with cell cycle arrest and apoptosis markers would elucidate if *gadd45b* upregulation is the cause of these effects in *sk8* embryos.

Growth arrest and DNA-damage inducible (GADD) family of proteins plays important roles in regulation of DNA repair, cell cycle control, and apoptosis (Mak and Kultz 2004). Expression of GADD45 family genes is induced by environmental stress (ultraviolet rays, high osmolarity) as well as by certain cytokines (interleukin 18, and 6) (Mak and Kultz 2004; Takekawa and Saito 1998). *Gadd45b* is regulated by NF-kappaB (De Smaele et al. 2001), and TGF-b (Takekawa et al. 2002) in mammalian cell cultures. NF-kB binding sites are detected in the promoter of *gadd45b* (Jin et al. 2002), suggesting direct transcriptional regulation by NF-kB members. At the functional level, Gadd45b protein binds to JNKK2 (c-Jun N-terminal kinase cascade member) directly and blocks its catalytic activity (Papa et al. 2004), thereby providing a molecular link between the NF-kB and JNK pathways in controlling apoptosis. Though we do not see differential expression of other downstream targets of NF-kB or TGF-beta in our array analysis of *sk8* embryos, we cannot exclude that possibility of specific transcriptional initiation of *gadd45b* by either or both of these pathways.

Future studies are required to elucidate if there is a mechanistic interaction between NF-kB/TGF-beta family members and Foggy/Spt5 to understand the interplay between transcription initiation and elongation regulation of *gadd45b*. It has been previously suggested that *gadd45* falls into the category of 'preset' genes whose upregulation is independent of changes in chromatin structure (by binding of transcription factors), or synthesis of new transcription factors, or stimuli-induced DNA-protein interactions (Graunke et al. 1999). Hence, it is tempting to speculate that the

elongation control of *gadd45b* by Foggy/Spt5 is the switch that regulates these 'preset' genes. Transcription elongation control of such 'preset' genes might be crucial to integrate signals from multiple pathways (*gadd45b* controlled by NF-kB and TGF-b) to coordinate regulation of important signaling pathways that ultimately determine cell fate and survival.

III. Gene expression profiling of *foggy* ^{m806} mutants

Background

The *foggy* mutant was identified in a genetic screen for mutations that affect neuronal development in zebrafish (Guo et al. 1999). In addition to multiple neuronal defects, the *foggy* mutant displayed a reduction in dopaminergic (DA) neurons and a corresponding increase in serotonergic (5-HT) neurons in the hypothalamus at 48 hours post fertilization (hpf) (Guo et al. 2000). Though the *foggy* mutant had relatively normal gross morphological features, it was lighter in color compared to wild-type (WT) embryos, developed circulatory problems by 36 hpf and died by 5 days post fertilization (dpf). Whole mount staining with multiple markers suggested that the *foggy* mutant neuronal phenotype might be due to defects in cell fate specification. Markers for cell death and proliferation were normal in *foggy* mutants.

Positional cloning identified the *foggy* gene to be homologous to *spt5*, a yeast transcription elongation factor. The *foggy* mutant allele carries a single nucleotide substitution from T to A, which changes the amino acid valine to aspartic acid (V1012D). Biochemical *in vitro* assays were performed to determine the effect of this mutation on the transcription elongation function of the protein. When recombinant wild-type (WT) Foggy protein was added to a transcription elongation complex with the drug 6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB), a nucleotide analog, there was a reduction in the rate of transcription elongation of an artificial construct. When the same assay was performed with mutant Foggy protein, there was no reduction in the rate of transcription elongation, suggesting the loss of the inhibitory activity of the mutant Foggy protein.

When recombinant WT Foggy protein was added to a transcription elongation complex with low nucleotide concentrations, elongation was stimulated. When the mutant Foggy protein was added to the complex under similar conditions, the elongation rate was comparable to WT levels, suggesting that the mutant Foggy protein had not lost its stimulatory activity. Combining *in vitro* biochemical data with the neuronal phenotype of the *foggy* mutant led to the hypothesis that the repressive activity of Foggy protein is essential for the development of distinct neurons in zebrafish. Preliminary studies of candidate genes known to be essential for DA and 5-HT neuronal development using whole mount *in situ* analysis did not yield promising results; most of these genes were unchanged in the *foggy* mutants. The goal of this study was to find target genes of Foggy/Spt5 that contribute to the observed neuronal phenotype.

Materials and Methods

Microarrays and qRT-PCR

Equal number of embryos (WT siblings, *foggy* mutants) at 48 hpf was dechorionated by hand or by Pronase K treatment. Heads of individual embryos were cut above the yolk using a needle and appropriate amounts of Trizol were then added to the embryos. Standard RNA extraction protocols were followed and total RNA was re-suspended in nuclease-free water. DNase treatment was performed to minimize contamination by genomic DNA. The total RNA sample was placed in dry ice and shipped to NINDS for amplification of mRNA and subsequent microarray hybridization.

To obtain cDNA, total RNA was reverse-transcribed using oligodT and Superscript enzyme. Equal amount of cDNA was used in SybrGreen qPCR mastermix system with appropriate primers. All conditions (primer concentration, and temperature for PCR) were optimized. All plates contained at least two well duplicates for each primer and cDNA condition. Cycle values were averaged between the duplicates and normalized to control primer values. WT fold values were compared to *foggy* mutant fold values and expressed as fold upregulated or downregulated in *foggy* embryos. At least three independent replicates were performed for each primer set.

Microarray Analysis

Microarray analysis was performed by Nathan Salomonis from Bruce Conklin's lab. Gene intensity values for the Affymetrix Zebrafish Genome Array CEL files were obtained using the GC-RMA package from Bioconductor (Dudoit et al. 2003; Wu et al. 2004). To identify genes with relative changes in steady-state mRNA levels between WT and *foggy* mutants, a fold change was calculated from these log2 intensity values along

with a student *ttest p* value. Genes with 2 fold change in expression and $p<0.05$ were clustered using the program HOPACH (hierarchical ordered partitioning and collapsing hybrid), with uncentered correlation distance (Pollard and van der Laan 2002; van der Laan and Pollard 2003).

To determine relative differences in gene length parameters, a custom python script was written that links microarray probe set data to gene structural information from the Ensembl database (PMID: 17148474) obtained through BioMart (Bioinformatics. 2005 Aug 15;21(16):3439-40). Probe set annotations to Ensembl genes obtained from BioMart were augmented with those directly from Affymetrix (http://www.affymetrix.com). In this script, fold changes were calculated by comparing the mean length of distinct gene features (genomic, exon and intron lengths/number) or gene expression between upregulated or downregulated Ensembl-linked genes compared to all genes on the microarray. For each parameter, a permutation *p* value was calculated by randomizing the selection of input probe sets 10,000 times and re-calculating and storing the mean fold changes for both up and down regulated sets to determine the probability that the original fold occurs by chance alone.

Whole mount staining:

Whole-mount *in situ* hybridization was performed as described before. Briefly, embryos were fixed in 4% paraformaldehyde, overnight in the cold room. Then, the embryos were dechorionated by hand and stored in 100% methanol at -20°C until usage. Embryos were rehydrated and treated mildly with proteinase K before incubation with appropriate digoxigenin-labeled RNA probes. After extensive washing at 60°C, an antibody against digoxigenin is added to the tubes and incubated overnight. NBT/BCIP

(Roche) colorimetric reaction is used to visualize the expression patterns of the genes. Whole-mount immunostaining was performed similarly with Anti-Parvalbumin antibody (1:10,000). Images were taken with Zeiss microscope.

Results and discussion:

Microarray data:

To identify genes that are differentially expressed between wild-type and *foggy*^{m806} mutant embryos, we performed microarray analysis by hybridizing zebrafish Affymetrix oligoarrays, which contain 15,617 probesets representing ~10,000 genes, with fluorescently labeled RNAs isolated from heads of *foggy* ^{m806} mutant embryos and their WT siblings at 48 hpf. Only 1.3% of the genes are differentially expressed (absolute fold greater than 2, $P < 0.05$) between *foggy* ^{m806} mutant embryos and WT siblings. Among the differentially expressed genes, 11 genes were upregulated (ranging from +2 to +5 fold), and 147 genes were downregulated (ranging from -2 to -28 fold) in *foggy* ^{m806} embryos. Most of the upregulated genes are novel, hypothetical proteins. Most of the downregulated genes are constituents of the extra-cellular matrix, calcium-binding proteins, and transporters.

To determine if differentially expressed genes in *foggy* ^{m806} embryos have specific gene structure requirements, we performed bioinformatics analysis on these gene sets (in collaboration with Nathan Salomonis, Bruce Conklin's lab). To assess the extent and overall likelihood of a difference between the length of Foggy/Spt5 regulated genomic structure to other genes in the array, we linked microarray probe sets to gene structure data from the Ensembl database, computed the mean of different gene features (e.g., genomic length) for differentially expressed genes and generated fold and P values using a permutation based analysis.

Categories	**Total**	**Up-regulated in *foggy* embryos**			**Down-regulated in *foggy* embryos**		
Groups	Mean	Mean	Fold	PermuteP	Mean	Fold	PermuteP
Gene length	18233	9353	-1.95	0.0458	13091	-1.39	0.012
Exon length	182	183	1.01	0.9146	179	-1.02	0.8216
Exon number	10	10	-1.05	0.0852	8	-1.26	0.8388
Intron length	2165	926	-2.34	0.0116	1849	-1.17	0.2338
Average expression	280	39	-7.19	0.77	415	1.48	0.1748
Genes	9955		11			147	
Probesets	15617		21			189	

Table 1: Summary of permutation based analysis of gene structure requirements of differentially expressed genes in *foggy* ^{m806} embryos. Upregulated genes have shorter gene length, shorter intron length and downregulated genes have shorter gene length compared to other genes on the array.

Interestingly, genes upregulated in *foggy* ^{m806} embryos were found to have on average ~1.95 fold decrease in overall gene length and ~2.3 fold decrease in intron length compared to all gene-linked probe sets on the array ($p<0.05$) (Table 1). In agreement with the permutation analysis from *fog* ^{sk8} mutant embryos, genes that are smaller in length and have shorter introns might be repressed by Foggy/Spt5 at the level of elongation. In contrast to differentially expressed genes in *fog* ^{sk8} mutant embryos, genes that are downregulated in *foggy* ^{m806} embryos were found to have on average ~1.39 fold decrease in overall gene length. It is likely that other characteristics (specific sequences, secondary structures in introns) play a role in determining the identities of targets of Foggy/Spt5.

			Microarray	qRT-PCR fold			
No.	**Probe ID**	**Genes**	**Description/ Function**	**Fold**	**48hpf Heads**	**48hpf-whole**	**36hpf-whole**	**30hpf-whole**
1	Dr.26372.1.A1_at	U-1	Unknown	3.21	6.25	5.8	3.93	1
2	Dr.15833.2.S1_at	U-2	Transcription	2.5	2.67	3.16	2.64	1
3	Dr.15991.1.S1_at	U-3	RING Finger protein	2.3	3.28	2.29	3.42	4.37
4	Dr.26333.1.A1_at	BRD2	Bromodomain protein	2.14	1.92	1.83	1	1
5	Dr.2022.1.A1_at	Arginase	Enzyme	2.07	2.26	1.39	1	1.91
6	Dr.8587.1.A2_at	IGFBP1	Metabolism	1.9	1.37	2.37	1.69	1
7	Dr.14282.1.S1_at	ATF3	Transcription factor	1.75	1.53	1.53	2.27	1.08
8	Dr.25657.1.A1_at	IER	Transcription factor	1.7	1.79	1.29	1	1.35
9	Dr.15033.1.S1_at	Xp8	Transcription factor	1.61	2.45	3.5	2.39	2.45
10	Dr.6208.1.A1_at	U-4	Unknown	1.5	2.29	1.93	2.42	1.92
11	Dr.9976.1.S1_at	Klf2B	Transcription factor	1.5	1.53	2.49	2.61	1.6
12	Dr.11266.1.S1_at	14-3-3	Protein-protein interaction	1.8	1.8	1.47	1	1
13	Dr.12804.1.S1_at	Somatostatin	Neuropeptide	2	3.1	2.87	2.53	1.95
14	Dr.3130.1.A1_at	Rgs4	G protein signaling	2.1	1.97	1.11	1	1
15	Dr.21244.1.S1_at	UCP2	Metabolism	2.1	2.69	1.38	1.15	1.18
16	Dr.3966.1.A1_at	NP25	Unknown	3.5	4.5	2.89	1.31	1.36
17	Dr.20928.1.S1_at	Parvalbumin 1d	Calcium-binding protein	4.12	5.3	1.16	1.18	1
18	Dr.5725.1.S1_at	HoxB6b	Transcription factor	4.5	2.41	1.72	2.34	1.64
19	Dr.2675.1.A1_at	FKBP5	Protein-protein interaction	6.7	14.7	4.29	10.80	3.3

Table 2: Quantitative RT-PCR verification of microarray data with RNA extracted from heads (48 hpf) of *foggy* ^{m806} and WT embryos. Red color represents genes that are upregulated in mutants and green color represents genes that are downregulated in mutants, compared to WT levels. qRTPCR was also performed from RNA extracted from whole embryos at earlier time points. Folds are averaged from at least three independent experiments.

Quantitative Real Time PCR (qRT-PCR)

As the mutant Foggy protein had lost its repressive activity but retained its stimulatory activity in *in vitro* assays (Guo et al. 2000), we primarily focused our experiments on upregulated genes from the microarray. Eleven upregulated genes, eight downregulated genes and two unchanged genes were further verified by qRT-PCR in WT and *foggy* ^{m806} embryos at 48 hpf. Differential expression was confirmed for all downregulated genes and 10/11 upregulated genes examined. A secondary list of 23 genes (18 novel genes) ranging from 1.5 to 2 fold was also picked for validation by qRT-PCR; 78% were verified at 48 hpf by this process. As the fold changes determined by qRTPCR did not show significant increases in fold, further analysis was not conducted on these genes.

The neuronal phenotype in *foggy* mutants is robust by 48 hpf. The genes that cause the phenotype are likely to have been mis-expressed at earlier time points. In order to separate genes that are causative of the phenotype from the genes that are the result of the phenotype, we performed qRTPCR with RNA extracted from WT and *foggy* mutant embryos at earlier time points (Table 2). Even though, RNA extracted from whole embryos were used at these earlier time points, eight out of eleven genes were upregulated at 36 hpf and four were upregulated as early as 30 hpf. Only two (FKBP5, somatostatin) out of eight downregulated genes were downregulated as early as 30 hpf. Interestingly, novel genes (U3, U4, p8) belong to this category of early-upregulated genes, suggesting that these genes might be causative of the neuronal phenotype in *foggy* mutants. Taken together, these results suggest two possibilities: these genes are highly mis-expressed in a restricted fashion (forebrain) in the whole mutant embryos (related to

the neuronal phenotype) or these genes are ubiquitously mis-expressed (not neuronal-specific) in the mutant embryos. Then, these genes might be more sensitive to the loss of repressive activity of the mutant Foggy protein and might be direct targets of Foggy/Spt5.

No.	ProbeID No.	Genes	*foggy*, 30 hpf	*sk8*, 29 hpf
1	Dr.15991.1.S1_at	U-3	4.37	16.51
2	Dr.26372.1.A1_at	U-1	1	16.09
3	Dr.15033.1.S1_at	Xp8	2.45	10.64
4	Dr.9976.1.S1_at	Klf2B	1.6	5.10
5	Dr.6208.1.A1_at	U-4	1.92	4.48
6	Dr.5725.1.S1_at	HoxB6b	1.64	4.05
7	Dr.15833.1.A1_at	U-2	1	3.51
8	Dr.26333.1.A1_at	BRD2	1	2.70
9	Dr.14282.1.S1_at	ATF3	1.08	2.40
10	Dr.3130.1.A1_at	Rgs4	1	2.17
11	Dr.25657.1.A1_at	IER	1.35	1.72
12	Dr.20928.1.S1_at	Parvalbumin 1d	1	15.60
13	Dr.12804.1.S1_at	Somatostatin	1.95	3.50
14	Dr.2022.1.A1_at	Arginase	1.91	2.23
15	Dr.21244.1.S1_at	UCP2	1.18	1.85
16	Dr.11266.1.S1_at	14-3-3	1	1.80

Table 3: qRT-PCR fold change comparison between *foggy* ^{m806} and *fog* ^{sk8} levels around 30 hpf. Red color represents genes that are upregulated in mutants and green color represents genes that are downregulated in mutants, compared to WT levels. Folds are averaged from at least three independent experiments.

In order to determine if these genes were direct targets of Foggy/Spt5, we performed qRT-PCR with the same genes on *fog* ^{sk8} mutant embryos at ~30hpf. *fog* ^{sk8} is a zygotic null allele of Foggy/Spt5, lacking both repressive and stimulatory activities. 9/11 upregulated genes and 4/8 downregulated genes in *foggy* is similarly mis-expressed in *sk8* embryos at 29 hpf (Table 3). Interestingly, four genes (U1, U3, Xp8, Parvalbumin 1d) show higher fold changes in *sk8* embryos than in *foggy* embryos, suggesting that the

in comparison to null allele (*fog* ^{sk8}), *foggy* ^{m806} is a partial loss of function allele. This observation is also consistent with the severe gross morphology defects of *sk8* embryos in comparison to the mild gross morphology defects of *foggy* mutant embryos. After narrowing down candidate genes from the microarray analysis by qRT-PCR validation and time course analysis, we performed further analysis on select set of genes in order to determine the spatial expression pattern and further infer possible function of these genes. Comprehensive results for some of the genes are summarized below.

Unknown 1 (U1):

ProbeID #: Dr.26372.1.A1_at

Unigene ID #: Dr. 122837 (as of 12/19/07)

Gene Structure: Unknown

Name of gene: Hypothetical protein XP_701190

Predicted function: Unknown

Results from this study:

Primers were designed according to the probe sequence on the microarrays towards the 3' end of the predicted gene sequence. The amplified product size was about 200 bp. U1 is expressed from 4 hpf to at least 72 hpf (Fig. 1a). This gene is not maternally expressed as there is no expression at 2 hpf. The RT-PCR gel also shows that the gene might be dynamically regulated based on the strength of the band at different time points. This gene is upregulated in *foggy* (36 – 48 hpf), and *sk8* (24 – 30 hpf) embryos. No *in situ* analysis or further follow-up was conducted due to lack of sequence information.

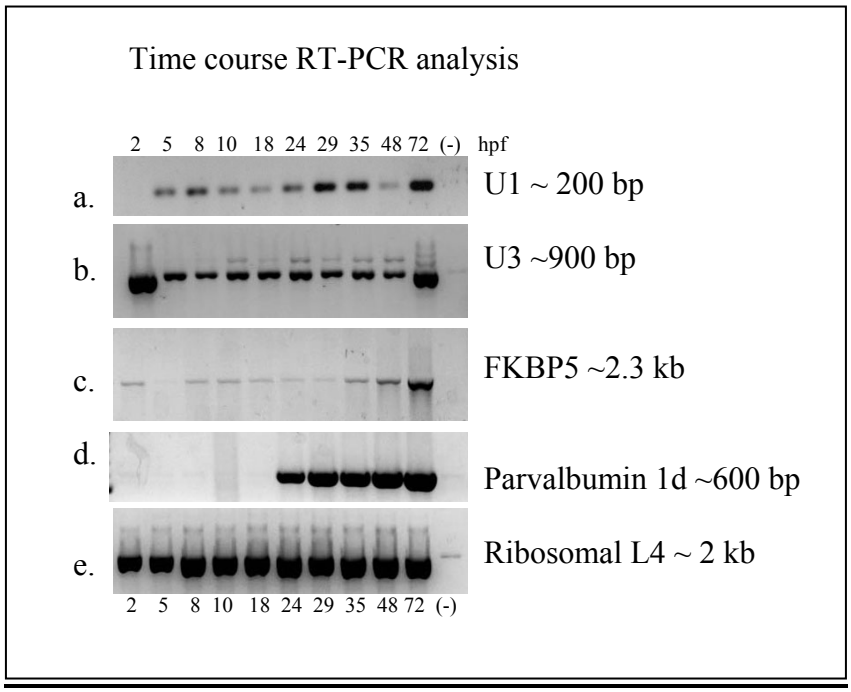


Figure 1: Time course RT-PCR analysis using primers in U1, U3, FKBP5, Parvalbumin 1d, Ribosomal L4 genes (control). (-) represent the (-) reverse transcriptase control column to determine specificity of primers and genomic DNA contamination in samples.

Unknown 3 (U3):

ProbeID #: Dr.15991.1.S1_at

Unigene ID #: Dr.83924 (as of 12/19/07)

Gene structure: Unknown

Name of gene: Hypothetical protein LOC100000332

Predicted function: Contains RING domain at N-terminus, possible E3 ubiquitin ligase

Results from this study:

U3 is expressed maternally and zygotic transcription starts at 10 hpf, when a 1.2 kb band appears along with the 900 bp band detected earlier (Fig. 1b). Expression

continues on to 72 hpf. This gene is upregulated in *foggy* (30-48hpf) and *sk8* (10-30 hpf) embryos. Using the predicted gene sequence from Ensembl, this gene was cloned from the zebrafish 24 hpf cDNA. U3 mRNA is weakly and ubiquitously expressed in WT embryos. U3 mRNA is highly and ubiquitously expressed in *sk8* embryos from 10-24 hpf. Over-expressing U3 mRNA into 1-8 cell stage of WT embryos did not produce any overt gross morphological phenotype from till 48 hpf. Immunostaining with Anti-TH (Tyrosine Hydroxylase) and Anti- 5HT (Serotonin) antibodies did not detect any significant differences between control injected and U3 injected embryos at 48 hpf.

FKBP5:

ProbeID #: Dr.2675.1.A1_at

Unigene ID #: Dr.78793 (as of 12/19/07)

Gene structure: Gene length - 19.87 kb, transcript length – 2.4 kb, no. of exons - 11

Name of gene: FK506 binding protein 5

Predicted function:

The protein encoded by this gene is a member of the immunophilin protein family, which play a role in immuno-regulation and basic cellular processes involving protein folding and trafficking. This encoded protein is a cis-trans prolyl isomerase that binds to the immuno-suppressants FK506 and rapamycin. It also interacts functionally with mature hetero-oligomeric progesterone receptor complexes along with the 90 kDa heat shock protein and P23 protein. Polymorphism in the *fkbp5* gene plays a role in the stress hormone-regulating hypothalamic-pituitary-adrenal axis. This polymorphism has

been found to be related to a faster response to antidepressant drug treatment and to increased recurrence of depressive episodes.

Results from this study:

FKBP5 mRNA is weakly expressed in during early stages of development (1 – 29 hpf). A 2.3 kb transcript is expressed robustly at 48 hpf and 72 hpf (Fig. 1c). *In situ* analysis shows ubiquitous expression in WT embryos, which is completely lost in *foggy* mutant embryos at 48 hpf. FKBP5 mRNA expression is not significantly affected in *sk8* embryos. ChIP with both anti-RNAPII and anti-Foggy antibodies in WT embryos show no occupancy on the gene at 24 hpf and considerable occupancy at 48 hpf. Further ChIP analysis in *foggy* mutant embryos at 48 hpf would determine if *fkbp5* is a direct target gene of Foggy/Spt5. It has been shown that yeast Ess1 and its mammalian homolog, Pin1 (Xu, 2007 #125; Xu, 2004 #126; Xu, 2003 #127; Kops, 2002 #128) interacts with phosphorylated form of Spt5 and opposes Spt4/5 (Lavoie, 2001 #129). It has been hypothesized that Ess1-induced conformational changes attenuate RNAPII elongation and help coordinate the ordered assembly of protein complexes on the CTD. Preliminary over-expression studies of *fkbp5* mRNA in WT embryos at 1-8 cell stage show minor migration defects in TH+ dopamine neurons at 48 hpf. Further study is required to determine if FKBP5 functions similarly with Foggy/Spt5 at the mechanistic level. It would also be interesting to determine if there is a causal connection between reduction of FKBP5 expression and the cell fate specification neuronal phenotype in *foggy* mutant embryos.

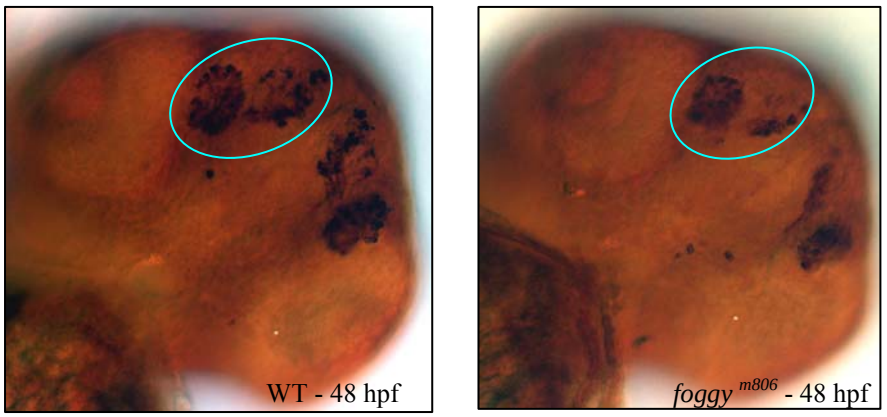


Figure 2: Whole mount immuno-staining of WT and *foggy* ^{m806} mutant embryos at 48 hpf with Anti-Parvalbumin antibody. Clusters of olfactory neurons are decreased in *foggy* mutants. Ventral view, 20X magnification.

Parvalbumin 1d:

ProbeID #: Dr.460.1.A1_at

Unigene ID #: Dr. (as of 12/19/07)

Gene structure: Gene length – 2.8 kb, transcript length – 711 bp, no. of exons - 5

Name of gene: Parvalbumin 1d

Predicted function: Calcium-binding protein

Results from this study:

Parvalbumin 1d expression starts at 24 hpf and is not maternally expressed (Fig. 1 d). Spatially, parvalbumin 1d mRNA is expressed in the somites. The parvalbumin gene family consists of nine members, six of which are downregulated in *foggy* and *sk8*

mutants (microarray data). Using mouse Anti-Parvalbumin antibody, we also detected clusters of neurons in the olfactory epithelium around 32 hpf. Interestingly, Anti-Parvalbumin antibody staining is reduced in specific clusters in *foggy* mutants at 48 hpf (Fig. 2). Time course analysis revealed that *foggy* mutant embryos displayed delayed development of Parvalbumin^{+} cells. In mouse, Parvalbumin^{+} cells in the olfactory epithelium are GABAergic neurons. Taken together, our results suggest that Parvalbumin^{+} GABAergic neurons are reduced in the forebrain of *foggy* mutants. Further studies are required to determine if the reduction is due to loss of progenitor populations at earlier time points (due to developmental delay) or death of these GABAergic cells.

Summary and future directions

Our gene expression profiling using RNA from heads of WT and *foggy* mutant embryos revealed that less than 1.5 % of genes are differentially expressed in the *foggy* mutants. Most of these genes are novel, uncharacterized genes. Initial characterization in this study has shown that most of these genes are ubiquitously expressed at earlier stages, suggesting essential roles during development. Permutation analysis of differentially expressed genes shows that genes that are shorter in gene length with shorter intron lengths are upregulated in *foggy* mutants. This observation is consistent with the *in vitro* data that shows that the mutant Foggy protein has lost its repressive activity, thereby leading to upregulation of its target genes. Interestingly, genes that are downregulated in *foggy* mutants also have shorter gene length compared to other genes on the array, suggesting that the mutant Foggy protein might be defective in some other activities.

As the zebrafish microarray used in this study contains ~10,000 protein coding

genes, factors required for DA/5-HT specification may not be all present in the array. None of the genes that are differentially expressed in *foggy* are known candidates of neuronal cell fate specification. The known factors essential for neuronal specification are not differentially expressed, at least at the steady-state mRNA level, in the mutant. Furthermore, the gene duplication event in the evolution of zebrafish and incomplete genomic sequence information (at the time of this study) made it difficult to determine the identities and functions of the candidate genes. Performing this study with microarrays that contain brain-specific factors might greatly enhance the profiling aspect of this study.

Though we were able to narrow down candidates based on mRNA mis-expression in *foggy* mutants at earlier time points by qRT-PCR, over-expression of these genes in WT embryos did not result in overt phenotypes. Reducing endogenous levels of novel genes (using morpholinos) identified in this study might be more effective at determining the function of these genes during development. It is also likely that the *foggy* mutant phenotype is caused by multiple genes and manipulating one gene at a time might not lead to a recapitulation of the mutant phenotype.

IV. Summary and Discussion

Foggy/Spt5 is a highly conserved transcription elongation factor. This protein has dual activities of repressing and stimulating transcript elongation. The association of Negative Elongation complex (NELF) and Foggy/Spt5 holds RNA Polymerase II (RNAPII) in a "pause" state, repressing elongation of the nascent transcript (Yamaguchi et al. 2002). This pausing mechanism allows the RNAPII complex to link early stages of transcription to mRNA processing on the chromatin, thereby ensuring efficiency and accuracy of transcription (Sims et al. 2004). In addition to this general processing functions required for most mRNA, pausing mechanism has also been studied as a specific and regulated step of gene expression for a handful of inducible genes (Saunders et al. 2006). By virtue of its association with multiple processing factors, Foggy/Spt5 also plays roles in stimulating transcription.

In vivo studies in worms, flies and zebrafish have provided considerable evidence to the physiological importance of Foggy/Spt5 in multi-cellular organisms (Cooper et al. 2005; Guo et al. 2000; Jennings et al. 2004; Keegan et al. 2002; Shim et al. 2002). Partial loss-of-function analyses in these various organisms suggest that Foggy/Spt5 is essential for proper differentiation of various cell types during development. Phenotypic mutant analysis also suggests that many different genes might be regulated by Foggy/Spt5. In spite of these initial characterizations of the *in vivo* physiological roles of Foggy/Spt5 in different model organisms, many questions remain. A).What genes are regulated by Foggy/Spt5? B).What type of elements (gene structure characteristics, function of the genes) defines Foggy/Spt5's target genes?

To address these questions, we have employed multiple molecular biology tools and zebrafish as our model organism. Most of the mechanistic studies on transcription elongation have been conducted in yeast and in cell culture. The standard assays involve depleting HeLa nuclear extracts with multiple antibodies to the general transcription machinery and then, adding-back purified recombinant proteins at different steps to determine their efficiency of transcribing an artificial DNA template (Wada et al. 1998). Though these studies have provided insights into the identities and functions of the individual factors involved in transcription elongation, the *in vivo* physiological relevance of transcription elongation factors and their targets is poorly understood.

Zebrafish is an ideal model organism to understand the role of transcription elongation factors in a vertebrate system. First, zebrafish development occurs externally making the early embryos amenable to manipulations and observations. Second, forward genetic screens in zebrafish have led to the identification of several mutant alleles of transcription elongation factors (*fog* ^{sk8}, *fog* ^{s30}, *foggy* ^{m806}, *foggy* ^{fh20}, *pandora, spt6 ko,* and *spt6* ^{sbu2}). Third, morpholino knock-down techniques have been successfully used in zebrafish to understand the roles of general transcription machinery in vertebrates (Bartfai et al. 2004; Ferg et al. 2007; Muller et al. 2001). Fourth, as early stages of zebrafish development occur mainly through maternal protein, it is highly likely that the manipulated embryos will be able to survive at least the initial stages of division, providing the researchers with enough analyzable biological material. Fifth, it is now possible to perform microarrays at different time points with multiple mutants, which greatly enhances the process of connecting the phenotype of the mutants to their gene expression profile.

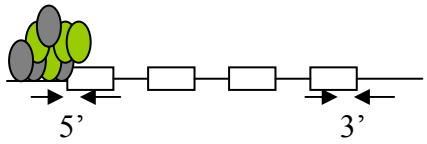
We have used two mutant alleles of *foggy* (*fog* ^{sk8}, *foggy* ^{m806}) in our study. The *fog* ^{sk8} allele is a zygotic null (Keegan et al. 2002), with mutant phenotype first detectable at 20 hours post fertilization (hpf). Most of the assays using this mutant were performed at 24 hpf when the mutant phenotype (gross developmental retardation) is robust. The other allele, *foggy* ^{m806} is considered a hypomorphic allele (Guo et al. 2000), with mutant phenotype first detectable around 30 hpf (slow heart beat, lighter color). As we were interested in the neuronal phenotype (reduction in dopamine neurons and increase in serotonin neurons in forebrain at 48 hpf) of this mutant, the microarray was performed using RNA from heads at 48 hpf.

Through expression profiling studies on ~10,000 protein-coding genes, a surprisingly small fraction (less than 5%) was differentially expressed between *fog*^{sk8} mutants and their wild-type (WT) siblings while less than 1.5% was differentially expressed between *foggy* ^{m806} and their WT siblings. Previous immunofluorescence analyses in *Drosophila* showed that Spt5 was localized to majority of sites on polytene chromosomes (Andrulis et al. 2000; Kaplan et al. 2000). These studies implied that most, if not all, genes may be dependent on Foggy/Spt5 for productive transcription. On the other hand, the ubiquitous localization and functional redundancy of transcription elongation factors *in vitro* implied that most genes might not be critically dependent on one factor for proper elongation. Our analyses do not exclude the possibility of larger number of Foggy/Spt5 target genes during other time points of development and/or with a more encompassing microarray. Further classification analyses revealed that these differentially expressed genes are involved in diverse biological pathways ranging from stress response to extra-cellular matrix components. Many previously uncharacterized,

novel genes were also differentially expressed in these mutants. Many of the unknown genes are expressed early in development and have ubiquitous *in situ* expression pattern. Further in-depth studies using loss-of-function morpholinos might illuminate the roles of these novel genes during vertebrate development.

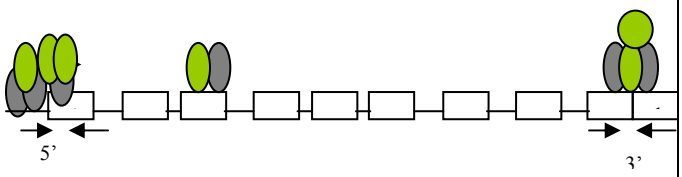
Repressive activity:

(1) Genes that are poised for transcription are occupied by both RNAPII and Foggy at only 5' end of the gene.
(2) These genes are shorter in gene length with shorter introns than average gene length.
(3) These genes are upregulated in *sk8, foggy* ^{m806} mutant embryos (lack of Foggy/Spt5).



Stimulatory activity:

(1) Genes that are actively transcribed are occupied by both RNAPII and Foggy at both 5' and 3' ends of the gene.
(2) These genes have average gene lengths.
(3) These genes are downregulated in *sk8* embryos.



Foggy might also be essential for stabilizing RNAPII on the chromatin

Figure 1: Classification of target genes of Foggy/Spt5. Grey ovals represent RNAPII occupancy and green ovals represent Foggy occupancy on the gene and rectangles represent exons.

A permutation based analysis revealed that genes upregulated in *fog*^{sk8} and *foggy*^{m806} embryos have significantly shorter gene length compared to the average gene length in zebrafish, suggesting an additional repressive level of regulation of shorter genes by Foggy/Spt5. One explanation for this observation is that the distinct advantage

of transcription elongation regulation is to allow rapid gene induction in response to extrinsic signals; genes that are regulated in this manner may have been selected during evolution to have shorter length, in order to further accommodate the need for immediate, inducible transcription. Further bioinformatics analysis using DNA sequence information of the target genes identified in this study might determine if specific sequences (primary or secondary structures) could act as binding sites for Foggy/Spt5.

Analysis of selected set of 21 genes (from the microarrays) in WT embryos using chromatin immunoprecipitation (ChIP) identified that a large fraction of these genes were occupied by Foggy/Spt5 in a pattern similar to RNAPII occupancy. Correlating RNAPII occupancy to transcriptional status of the gene in WT embryos, we found that many of the upregulated genes were poised for transcription while many downregulated genes were actively transcribed. This observation suggests that Foggy/Spt5 acts as a repressor for many of the poised genes and as a stimulator for many of the actively transcribed genes identified in this study (Fig. 1).

Spt4-5 complex is important for stabilizing RNAPII on the chromatin template under normal and DNA-damage inducing conditions in yeast (Mason and Struhl 2005), (Jansen et al. 2002). Our RNAPII ChIP analysis in sk8 embryos suggest that Foggy/Spt5 is important for stabilizing RNAPII on the chromatin of vertebrates as well, suggesting that RNAPII might be considered a target of Foggy/Spt5. Analysis using antibodies recognizing different conformations of RNAPII would further elucidate the mechanism by which Foggy/Spt5 stabilizes particular conformations of RNAPII.

Our comparison of RNAPII occupancy between WT and *sk8* embryos identified Growth Arrest and DNA-damage inducible gene (*gadd45b*) to be a novel target for

repression by Foggy. Growth arrest and DNA-damage inducible (GADD) family of proteins plays important roles in regulation of DNA repair, cell cycle control, apoptosis and development (De Smaele et al. 2001; Kawahara et al. 2005; Larsen et al. 2006; Mak and Kultz 2004; Papa et al. 2007; Papa et al. 2004; Peretz et al. 2007; Takekawa and Saito 1998; Takekawa et al. 2002). Expression of GADD45 family genes is induced by environmental stress (ultraviolet rays, high osmolarity) as well as by certain cytokines (interleukin 18, and 6). *Gadd45b* is regulated by NF-kappaB (De Smaele et al. 2001), and TGF-b (Takekawa et al. 2002) in mammalian cell cultures. In zebrafish, overexpression of *gadd45b* by mRNA injection at the 1-cell stage leads to somite segmentation defects (down-regulation of *myoD*) (Kawahara et al. 2005). As *gadd45b* is not overexpressed in *sk8* embryos until 24 hpf, it is interesting to note that *sk8* embryos do not show obvious segmentation defects and no reduction in *myoD* transcript levels, suggesting that temporal and spatial context greatly influence the phenotypic outcome of *gadd45b* overexpression during development.

It has been previously suggested that *gadd45* falls into the category of 'preset' genes whose upregulation is independent of changes in chromatin structure (by binding of transcription factors), or synthesis of new transcription factors, or stimuli-induced DNA-protein interactions (Graunke et al. 1999). Hence, it is tempting to speculate that the elongation control of *gadd45b* by Foggy/Spt5 is the switch that regulates these 'preset' genes. Transcription elongation control of such 'preset' genes might be crucial to integrate signals from multiple pathways (e.g., *gadd45b* controlled by NF-kB and TGF-b pathways) to coordinate regulation of important signaling pathways that ultimately determine cell fate and survival (Fig. 2).

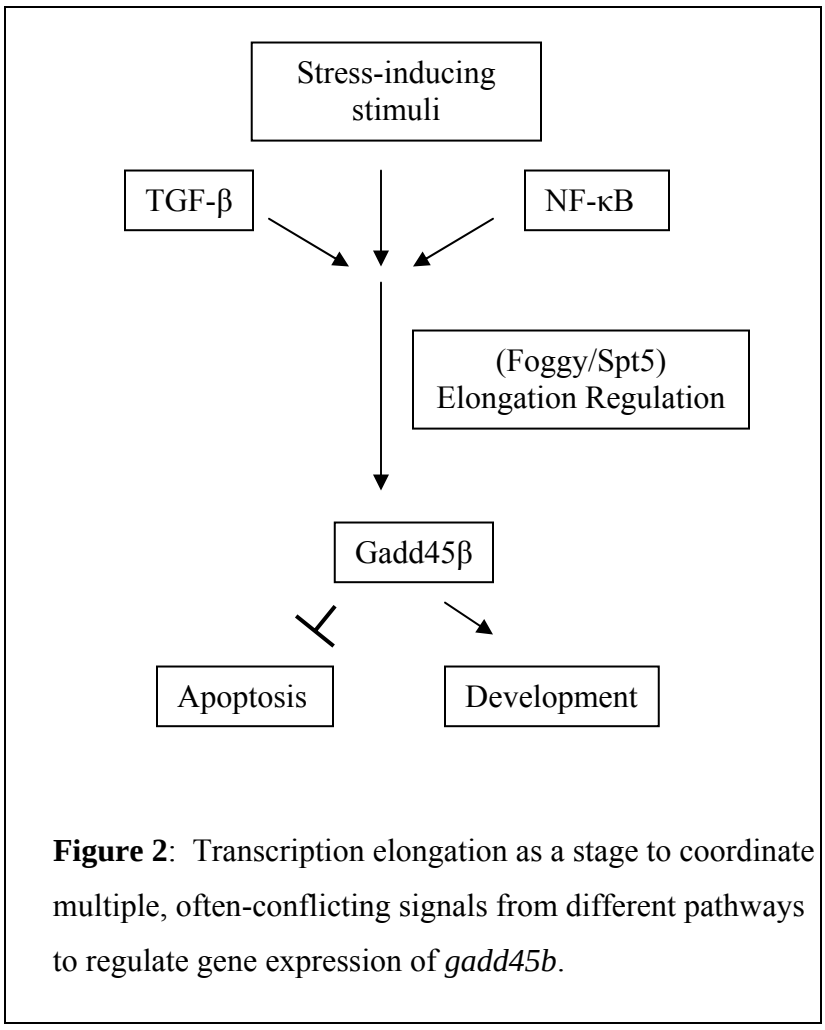


Figure 2: Transcription elongation as a stage to coordinate multiple, often-conflicting signals from different pathways to regulate gene expression of *gadd45b*.

Even though the general patterns of RNAPII and Foggy occupancies have been identified with a small number of genes from the microarrays, in-depth analysis with more genes at multiple time points will reveal intricacies of transcriptional regulation. Future work involving traditional elongation assays would be required to identify the exact nucleotide positions of pausing on these genes. Such studies in the future with more number of genes will enable the community to answer questions regarding the structural characteristics of the pause state of RNAPII complex.

From the development point of view, future work on determining the identities and functions of novel genes identified in this study will determine how these genes affect the mutants' phenotype. This would be particularly exciting for the neuronal cell fate specification phenotype of the *foggy* mutant. It is likely that the mutants' phenotype is caused by multiple genes with diverse functions; hence, GenMAPP and Panther pathway analysis (Dahlquist et al. 2002; Salomonis et al. 2007) would be indispensable tools to map genes into pathways and to understand the network of the genes in relation to the phenotype.

In conclusion, our study provides interesting insights into the identities and characteristics of target genes of Foggy/Spt5 in vertebrates and further studies are required to illuminate the role of these genes during development.

V. References

Andrulis, E. D., E. Guzman, P. Doring, J. Werner and J. T. Lis (2000). "High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation." Genes Dev **14**(20): 2635-49.

Bartfai, R., C. Balduf, T. Hilton, Y. Rathmann, Y. Hadzhiev, L. Tora, L. Orban and F. Muller (2004). "TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish." Curr Biol **14**(7): 593-8.

Bortvin, A. and F. Winston (1996). "Evidence that Spt6 controls chromatin structure by a direct interaction with histone." Science **272**: 1473-1476.

Bourgeois, C. F., Y. K. Kim, M. J. Churcher, M. J. West and J. Karn (2002). "Spt5 cooperates with human immunodeficiency virus type 1 Tat by preventing premature RNA release at terminator sequences." Mol Cell Biol **22**(4): 1079-93.

Bregman, D. B., L. Du, Y. Li, S. Ribisi and S. L. Warren (1994). "Cytostellin distributes to nuclear regions enriched with splicing factors." J Cell Sci **107 (Pt 3)**: 387-96.

Chiang, P. W., S. Q. Wang, P. Smithivas, W. J. Song, E. Crombez, A. Akhtar, R. Im, J. Greenfield, S. Ramamoorthy, M. Van Keuren, C. C. Blackburn, C. H. Tsai and D. M. Kurnit (1996). "Isolation and characterization of the human and mouse homologues (SUPT4H and Supt4h) of the yeast SPT4 gene." Genomics **34**(3): 368-75.

Chodosh, L. A., A. Fire, M. Samuels and P. A. Sharp (1989). "5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole inhibits transcription elongation by RNA polymerase II in vitro." J Biol Chem **264**(4): 2250-7.

Collart, M. A., N. Tourkine, D. Belin, P. Vassalli, P. Jeanteur and J. M. Blanchard (1991). " c-fos gene transcription in murine macrophages is modulated by a calcium-dependent block to elongation in intron 1." Mol. Cell Biol. **11**: 2826-31.

Compagnone-Post, P. A. and M. A. Osley (1996). "Mutations in the SPT4, SPT5, and SPT6 genes alter transcription of a subset of histone genes in Saccharomyces cerevisiae." Genetics **143**: 1543-1554.

Cooper, K. L., J. Armstrong and C. B. Moens (2005). "Zebrafish foggy/spt 5 is required for migration of facial branchiomotor neurons but not for their survival." Dev Dyn **234**(3): 651-8.

Dahlquist, K. D., N. Salomonis, K. Vranizan, S. C. Lawlor and B. R. Conklin (2002). "GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways." Nat Genet **31**(1): 19-20.

De Smaele, E., F. Zazzeroni, S. Papa, D. U. Nguyen, R. Jin, J. Jones, R. Cong and G. Franzoso (2001). "Induction of gadd45beta by NF-kappaB downregulates pro-apoptotic JNK signalling." Nature **414**(6861): 308-13.

Ferg, M., R. Sanges, J. Gehrig, J. Kiss, M. Bauer, A. Lovas, M. Szabo, L. Yang, U. Straehle, M. J. Pankratz, F. Olasz, E. Stupka and F. Muller (2007). "The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish." Embo J **26**(17): 3945-56.

Fivaz, J., M. C. Bassi, S. Pinaud and J. Mirkovitch (2000). "RNA polymerase II promoter-proximal pausing upregulates c-fos gene expression." Gene **255**(2): 185-94.

Fong, Y. W. and Q. Zhou (2001). "Stimulatory effect of splicing factors on transcriptional elongation." Nature **414**(6866): 929-33.

Furger, A., J. M. O'Sullivan, A. Binnie, B. A. Lee and N. J. Proudfoot (2002). "Promoter proximal splice sites enhance transcription." Genes Dev **16**(21): 2792-9.

Giardina, C. and J. T. Lis (1993). "Polymerase processivity and termination on Drosophila heat shock genes." J Biol Chem **268**(32): 23806-11.

Gilmour, D. S. and J. T. Lis (1986). "RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells." Mol. Cell Biol. **6**: 3984-3989.

Graunke, D. M., A. J. Fornace, Jr. and R. O. Pieper (1999). "Presetting of chromatin structure and transcription factor binding poise the human GADD45 gene for rapid transcriptional up-regulation." Nucleic Acids Res **27**(19): 3881-90.

Guenther, M. G., S. S. Levine, L. A. Boyer, R. Jaenisch and R. A. Young (2007). "A chromatin landmark and transcription initiation at most promoters in human cells." Cell **130**: 77-88.

Guo, S., Y. Yamaguchi, S. Schilbach, T. Wada, A. Goddard, J. Lee, D. French, H. Handa and A. Rosenthal (2000). "A regulator of transcriptional elongation, which is required for vertebrate neuronal development." Nature **408**: 366-369.

Happel, A. M., M. S. Swanson and F. Winston (1991). "The SNF2, SNF5 and SNF6 genes are required for Ty transcription in Saccharomyces cerevisiae." Genetics **128**(1): 69-77.

Hartzog, G. A., T. Wada, H. Handa and F. Winston (1998). "Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae." Genes Dev **12**(3): 357-69.

Ivanov, D., Y. T. Kwak, J. Guo and R. B. Gaynor (2000). "Domains in the SPT5 protein that modulate its transcriptional regulatory properties." Mol Cell Biol **20**(9): 2970-83.

Jansen, L. E., A. I. Belo, R. Hulsker and J. Brouwer (2002). "Transcription elongation factor Spt4 mediates loss of phosphorylated RNA polymerase II transcription in response to DNA damage." Nucleic Acids Res **30**(16): 3532-9.

Jennings, B. H., S. Shah, Y. Yamaguchi, M. Seki, R. G. Phillips, H. Handa and D. Ish-Horowicz (2004). "Locus-specific requirements for Spt5 in transcriptional activation and repression in Drosophila." Curr Biol **14**(18): 1680-4.

Jin, R., E. De Smaele, F. Zazzeroni, D. U. Nguyen, S. Papa, J. Jones, C. Cox, C. Gelinas and G. Franzoso (2002). "Regulation of the gadd45beta promoter by NF-kappaB." DNA Cell Biol. **21**: 491-503.

Jones, J. C., H. P. Phatnani, T. A. Haystead, J. A. MacDonald, S. M. Alam and A. L. Greenleaf (2004). "C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats." J Biol Chem **279**(24): 24957-64.

Kaplan, C. D., J. R. Morris, C. Wu and F. Winston (2000). "Spt5 and spt6 are associated with active transcription and have characteristics of general elongation factors in D. melanogaster." Genes Dev **14**(20): 2623-34.

Kawahara, A., Y. S. Che, R. Hanaoka, H. Takeda and I. B. Dawid (2005). "Zebrafish GADD45beta genes are involved in somite segmentation." Proc Natl Acad Sci U S A **102**(2): 361-6.

Keegan, B. R., J. L. Feldman, D. H. Lee, D. S. Koos, R. K. Ho, D. Y. Stainier and Y. D. (2002). "The elongation factors Pandora/Spt6 and Foggy/Spt5 promote transcription in the zebrafish embryo." Development **129**: 1623-1632.

Kim, M., S. H. Ahn, N. J. Krogan, J. F. Greenblatt and S. Buratowski (2004). "Transitions in RNA polymerase II elongation complexes at the 3' ends of genes." Embo J **23**(2): 354-64.

Kim, T. H., L. O. Barrera, M. Zheng, C. Qu, M. A. Singer, T. A. Richmond, Y. Wu, R. D. Green and B. Ren (2005). "A high-resolution map of active promoters in the human genome." Nature **436**: 876-880.

Kornblihtt, A. R., M. de la Mata, J. P. Fededa, M. J. Munoz and G. Nogues (2004). "Multiple links between transcription and splicing." Rna **10**(10): 1489-98.

Kwak, Y. T., J. Guo, S. Prajapati, K. J. Park, R. M. Surabhi, B. Miller, P. Gehrig and R. B. Gaynor (2003). "Methylation of SPT5 regulates its interaction with RNA polymerase II and transcriptional elongation properties." Mol Cell **11**(4): 1055-66.

Kyrpides, N. C., C. R. Woese and C. A. Ouzounis (1996). "KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins." Trends Biochem Sci **21**(11): 425-6.

Lanchy, J. M., C. Isel, C. Ehresmann, R. Marquet and B. Ehresmann (1996). "Structural and functional evidence that initiation and elongation of HIV-1 reverse transcription are distinct processes." Biochimie **78**(11-12): 1087-96.

Larsen, C. M., M. G. Dossing, S. Papa, G. Franzoso, N. Billestrup and T. Mandrup-Poulsen (2006). "Growth arrest- and DNA-damage-inducible 45beta gene inhibits c-Jun N-terminal kinase and extracellular signal-regulated kinase and decreases IL-1beta-induced apoptosis in insulin-producing INS-1E cells." Diabetologia **49**(5): 980-9.

Li, B., J. A. Weber, Y. Chen, A. L. Greenleaf and D. S. Gilmour (1996). "Analyses of promoter-proximal pausing by RNA polymerase II on the hsp70 heat shock gene promoter in a Drosophila nuclear extract." Mol Cell Biol **16**(10): 5433-43.

Mak, S. K. and D. Kultz (2004). "Gadd45 proteins induce G2/M arrest and modulate apoptosis in kidney cells exposed to hyperosmotic stress." J Biol Chem **279**(37): 39075-84.

Mandal, S. S., C. Chu, T. Wada, H. Handa, A. J. Shatkin and D. Reinberg (2004). "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 **101**(20): 7572-7.

Mason, P. B. and K. Struhl (2005). "Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo." Mol Cell **17**(6): 831-40.

Mi, H., N. Guo, A. Kejariwal and P. D. Thomas (2007). "PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways." Nucleic Acids Res **35**(Database issue): D247-52.

Mikkelsen, T. S., M. Ku, D. B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T. K. Kim, R. P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum,

E. S. Lander and B. E. Bernstein (2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." Nature **448**: 553-560.
Morris, D. P., H. P. Phatnani and A. L. Greenleaf (1999). "Phospho-carboxyl-terminal domain binding and the role of a prolyl isomerase in pre-mRNA 3'-End formation." J Biol Chem **274**(44): 31583-7.
Muller, F., L. Lakatos, J. Dantonel, U. Strahle and L. Tora (2001). "TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish." Curr Biol **11**(4): 282-7.
Myers, J. K., D. P. Morris, A. L. Greenleaf and T. G. Oas (2001). "Phosphorylation of RNA polymerase II CTD fragments results in tight binding to the WW domain from the yeast prolyl isomerase Ess1." Biochemistry **40**(29): 8479-86.
Ni, Z., B. E. Schwartz, J. Werner, J. R. Suarez and J. T. Lis (2004). "Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes." Mol Cell **13**(1): 55-65.
O'Brien, T. and J. T. Lis (1991). "RNA polymerase II pauses at the 5' end of the transcriptionally induced Drosophila hsp70 gene." Mol Cell Biol **11**(10): 5285-90.
Orphanides, G. and D. Reinberg (2002). "A unified theory of gene expression." Cell **108**(4): 439-51.
Papa, S., S. M. Monti, R. M. Vitale, C. Bubici, S. Jayawardena, K. Alvarez, E. De Smaele, N. Dathan, C. Pedone, M. Ruvo and G. Franzoso (2007). "Insights into the structural basis of the GADD45beta-mediated inactivation of the JNK kinase, MKK7/JNKK2." J Biol Chem **282**(26): 19029-41.
Papa, S., F. Zazzeroni, C. Bubici, S. Jayawardena, K. Alvarez, S. Matsuda, D. U. Nguyen, C. G. Pham, A. H. Nelsbach, T. Melis, E. De Smaele, W. J. Tang, L. D'Adamio and G. Franzoso (2004). "Gadd45 beta mediates the NF-kappa B suppression of JNK signalling by targeting MKK7/JNKK2." Nat Cell Biol **6**(2): 146-53.
Pei, Y. and S. Shuman (2002). "Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5." J Biol Chem **277**(22): 19639-48.
Peretz, G., A. Bakhrat and U. Abdu (2007). "Expression of the Drosophila melanogaster GADD45 Homolog (CG11086) Affects Egg Asymmetric Development That Is Mediated by the c-Jun N-Terminal Kinase Pathway." Genetics **177**(3): 1691-702.
Pinaud, S. and J. Mirkovitch (1998). "Regulation of c-fos expression by RNA polymerase elongation competence." J Mol Biol **280**(5): 785-98.
Plet, A., D. Eick and J. M. Blanchard (1995). "Elongation and premature termination of transcripts initiated from c-fos and c-myc promoters show dissimilar patterns." Oncogene **10**(2): 319-28.
Ponting, C. P. (2002). "Novel domains and orthologues of eukaryotic transcription elongation factors." Nucleic Acids Res **30**(17): 3643-52.
Rasmussen, E. B. and J. T. Lis (1993). "In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes." Proc Natl Acad Sci U S A **90**(17): 7923-7.
Rougvie, A. E. and J. T. Lis (1988). "The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged." Cell **54**(6): 795-804.
Ryser, S., T. Fujita, S. Tortola, I. Piuz and W. Schlegel (2007). "The rate of c-fos transcription in vivo is continuously regulated at the level of elongation by

dynamic stimulus-coupled recruitment of positive transcription elongation factor b." J Biol Chem **282**(7): 5075-84.
Salomonis, N., K. Hanspers, A. C. Zambon, K. Vranizan, S. C. Lawlor, K. D. Dahlquist, S. W. Doniger, J. Stuart, B. R. Conklin and A. R. Pico (2007). "GenMAPP 2: new features and resources for pathway analysis." BMC Bioinformatics **8**: 217.
Saunders, A., L. J. Core and J. T. Lis (2006). "Breaking barriers to transcription elongation." Nat Rev Mol Cell Biol **7**(8): 557-67.
Saunders, A., J. Werner, E. D. Andrulis, T. Nakayama, S. Hirose, D. Reinberg and J. T. Lis (2003). "Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo." Science **301**: 1094-6.
Shaw, P. E. (2007). "Peptidyl-prolyl cis/trans isomerases and transcription: is there a twist in the tail?" EMBO Rep **8**(1): 40-5.
Shim, E. Y., A. K. Walker, Y. Shi and T. K. Blackwell (2002). "CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the C. elegans embryo." Genes Dev **16**(16): 2135-46.
Sims, R. J., 3rd, R. Belotserkovskaya and D. Reinberg (2004). "Elongation by RNA polymerase II: the short and long of it." Genes Dev **18**(20): 2437-68.
Somesh, B. P., J. Reid, W. F. Liu, T. M. Sogaard, H. Erdjument-Bromage, P. Tempst and J. Q. Svejstrup (2005). "Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest." Cell **121**(6): 913-23.
Swanson, M. S., E. A. Malone and F. Winston (1991). "SPT5, an essential gene important for normal transcription in Saccharomyces cerevisiae, encodes an acidic nuclear protein with a carboxy-terminal repeat." Mol Cell Biol **11**(8): 4286.
Swanson, M. S. and F. Winston (1992). "SPT4, SPT5 and SPT6 interactions: effects on transcription and viability in *Saccharomyces cerevisiae*." Genetics **132**: 325-336.
Takekawa, M. and H. Saito (1998). "A family of stress-inducible GADD45-like proteins mediate activation of the stress-responsive MTK1/MEKK4 MAPKKK." Cell **95**(4): 521-30.
Takekawa, M., K. Tatebayashi, F. Itoh, M. Adachi, K. Imai and H. Saito (2002). "Smad-dependent GADD45beta expression mediates delayed activation of p38 MAP kinase by TGF-beta." Embo J **21**(23): 6473-82.
Thisse, B., S. Pfumio, M. Fürthauer, L. B., V. Heyer, A. Degrave, R. Woehl, A. Lux, T. Steffan, X. Q. Charbonnier and C. Thisse (2004). "Expression of the zebrafish genome during embryogenesis." ZFIN on-line publication.
Uptain, S. M., C. M. Kane and M. J. Chamberlin (1997). "Basic mechanisms of transcription elongation and its regulation." Annu. Rev. Biochem. **66**: 117-172.
van Haasteren, G., S. Li, S. Ryser and W. Schlegel (2000). "Essential contribution of intron sequences to Ca(2+)-dependent activation of c-fos transcription in pituitary cells." Neuroendocrinology **72**(6): 368-78.
Wada, T., T. Takagi, Y. Yamaguchi, A. Ferdous, T. Imai, S. Hirose, S. Sugimoto, K. Yano, G. A. Hartzog, F. Winston, S. Buratowski and H. Handa (1998). "DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs." Genes Dev. **12**: 343-356.

Wilcox, C. B., A. Rossettini and S. D. Hanes (2004). "Genetic interactions with C-terminal domain (CTD) kinases and the CTD of RNA Pol II suggest a role for ESS1 in transcription initiation and elongation in Saccharomyces cerevisiae." Genetics **167**(1): 93-105.

Winston, F., D. T. Chaleff, B. Valent and G. R. Fink (1984). "Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae." Genetics **107**: 179-197.

Winston, F. and P. Sudarsanam (1998). " The SAGA of Spt proteins and transcriptional analysis in yeast: past, present, and future." Cold Spring Harb Symp Quant Biol. **63**: 553-61.

Wu-Baer, F., W. S. Lane and R. B. Gaynor (1998). "Role of the human homolog of the yeast transcription factor SPT5 in HIV-1 Tat-activation." J Mol Biol **277**(2): 179-97.

Wu, C. H., C. Lee, R. Fan, M. J. Smith, Y. Yamaguchi, H. Handa and D. S. Gilmour (2005). "Molecular characterization of Drosophila NELF." Nucleic Acids Res **33**(4): 1269-79.

Wu, C. H., Y. Yamaguchi, L. R. Benjamin, M. Horvat-Gordon, J. Washinsky, E. Enerly, J. Larsson, A. Lambertsson, H. Handa and D. Gilmour (2003). "NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila." Genes Dev **17**(11): 1402-14.

Wu, X., A. Rossettini and S. D. Hanes (2003). "The ESS1 prolyl isomerase and its suppressor BYE1 interact with RNA pol II to inhibit transcription elongation in Saccharomyces cerevisiae." Genetics **165**(4): 1687-702.

Wu, X., C. B. Wilcox, G. Devasahayam, R. L. Hackett, M. Arevalo-Rodriguez, M. E. Cardenas, J. Heitman and S. D. Hanes (2000). "The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery." Embo J **19**(14): 3727-38.

Xiao, Y., Y. H. Yang, T. A. Burckin, L. Shiue, G. A. Hartzog and M. R. Segal (2005). "Analysis of a splice array experiment elucidates roles of chromatin elongation factor Spt4-5 in splicing." PLoS Comput Biol **1**(4): e39.

Yamada, T., Y. Yamaguchi, N. Inukai, S. Okamoto, T. Mura and H. Handa (2006). "P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation." Mol Cell **21**(2): 227-37.

Yamaguchi, Y., N. Inukai, T. Narita, T. Wada and H. Handa (2002). "Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA." Mol Cell Biol **22**(9): 2918-27.

Yamaguchi, Y., T. Wada, D. Watanabe, T. Takagi, J. Hasegawa and H. Handa (1999). "Structure and function of the human transcription elongation factor DSIF." J Biol Chem **274**(12): 8085-92.

VI. Library release form

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Keerthi Krishnan — Author Signature

1/2/18 — Date