Open Research Online

The Open University's repository of research publications
and other research outputs

The Super Elongation Complex (SEC) in Development and Disease

Thesis

How to cite:

Lin, Chengqi (2013). The Super Elongation Complex (SEC) in Development and Disease. PhD thesis The Open University.

For guidance on citations see FAQs.

© 2013 The Author

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data policy on reuse of materials please consult the policies page.

oro.open.ac.uk

UNRESTRICTED

The Super Elongation Complex

(SEC)

in Development and Disease

A Thesis Submitted for the Degree of

Doctor of Philosophy

by

Chengqi Lin

The Stowers Institute for Medical Research,

an Affiliated Research Centre of the Open University

August 2012

DATE OF SUBMISSION: 11 SEPTEMBER 2012

DATE OF AWARD: 29 JANUARY 2013

ProQuest Number: 13835920

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

ProQuest 13835920

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Acknowledgement

The research presented in this dissertation was conducted at the Stowers Institute for Medical Research and was supported by the National Institute of Health and the Stowers Institute for Medical Research.

I am extremely grateful to my supervisor, Dr. Ali Shilatifard, for his constant support, invaluable advice, and encouragement throughout my studentship. His support and understanding are deeply appreciated. I would also like to extend my thanks to all members of the Shilatifard laboratory: Edwin and Zhuojuan for the collaboration and thoughtful discussions; Alex for bioinformatics support; Pengfei, Min, and Stacy for the helpful experimental support; Kai-Chun and Eliza for antibody generation and characterization; Man, Nima, Shima, Jun-Shin, Yo-hei, and Hans-Martin for their assistance during the study; Laura for help with the editing of manuscripts; Lisa for ensuring everything ran smoothly.

I am also grateful to my thesis committee members Robb Krumlauf, Paul Trainor, Ron Conaway, Linheng Li, and Jerry Workman for their expertise and valuable suggestions. I would like to thank Robb Krumlauf and members of the Krumlauf laboratory, Bony and Mark, for sharing their microarray data from retinoic acid-induced ES cells. I would like to thank Tari, Valerie, Maria, Alexis, and David from the Tissue Culture Center for their support. I would also like to thank Michael, Laurence, and Selene from the Proteomics Center for their analyses of the MudPIT data.

Finally, I thank my parents, sister, brother, and parents-in-law for their love, moral support and encouragement. I would especially like to thank my wife, Zhuojuan, and my daughter, Sarah, for their love and continuous support.

Abstract

Chromosomal translocations involving the *mixed lineage leukemia* (*MLL*) gene are associated with infant acute leukemia. There are a large number of translocation partners of MLL that share very little sequence similarities, yet their translocations into MLL result in the pathogenesis of leukemia. To define the molecular reason why these translocations result in leukemogenesis, I purified several of the commonly occurring MLL chimeras and identified a novel Super Elongation Complex (SEC) associated with all chimeras purified. SEC consists of the RNA Pol II elongation factors ELL1-3, P-TEFb, and several frequent MLL-translocation partners. SEC is one of the most active P-TEFb complexes and is required for the proper expression of MLL chimera target genes and the oncogene, *MYC*, suggesting that the regulation of transcription elongation checkpoint control (TECC) by SEC could play essential roles in leukemia.

Paused Pol II has been proposed to be associated with loci that respond rapidly to environmental stimuli. My studies in mouse ES cells demonstrated that SEC is required for rapid transcriptional activation of genes, many of which contain paused Pol II. However, SEC is also required for the activation of the *Cyp26a1* gene, which does not contain detectable Pol II, yet responds much more rapidly to retinoic acid than those paused genes, suggesting that paused Pol II is not a prerequisite for rapid gene activation. Furthermore, Ell3, a member of the ELL family of proteins, predominately occupies poised, active, and inactive enhancers of many developmental genes in ES cells. Ell3's association with enhancers is required for setting up proper Pol II occupancy at the promoter-proximal regions of neighboring genes, providing a yet to be discovered mechanism for the transition from Ell3's presence at poised enhancers in ES cells to Ell2's role in the release of paused Pol II during gene activation.

Table of Contents

Acknowledgement II

Abstract III

Table of Contents IV

Table of Figures IX

Table of Tables XII

Abbreviations XII

Chapter 1. Introduction 1

1.1. Transcriptional elongation control by RNA polymerase II 1

1.1.1. RNA polymerases in mammals 1

1.1.1.1. The identification of three mammalian RNA polymerases 1

1.1.1.2. The structure of RNA polymerase II 3

1.1.1.3. RNA polymerase II large subunit Carboxyl Terminal Domain 5

1.1.2. Transcription by RNA polymerase II 8

1.1.2.1. The assembly of the pre-initiation complex 8

1.1.2.2. Early elongation 11

1.1.2.3. Promoter-proximal pausing 12

The establishment of RNA Pol II pausing 14

Factors involved in reducing transient pausing 16

1.1.2.4. P-TEFb and productive elongation 22

1.1.2.5. Transcriptional elongation-coupled RNA processing 24

RNA capping 25

Pre-mRNA splicing 25

3' end processing 26

1.2. MLL rearrangement-mediated leukemogenesis 28

1.2.1. *MLL* encodes a histone H3 lysine 4 methyltransferase 29

1.2.1.1. *MLL* is a homolog of the *Drosophila trithorax* gene 29

1.2.1.2. MLL forms a COMPASS-like complex 32

1.2.2. Characteristics of MLL-translocated leukemia 33
1.2.2.1. 11q23 rearrangements 33
1.2.2.2. MLL chimera lacks the conserved enzymatic SET domain 34
1.2.2.3. Gene expression signature 35
1.2.2.4. The transcriptome of MLL LSC is more akin to ESC than HSC 36
1.2.3. Diverse mechanisms for MLL chimera-mediated gene expression 37
1.2.3.1. Transcriptional Elongation Checkpoint Control (TECC) in leukemia 38
1.2.3.2. DOT1L-mediated H3K79 methylation 39
1.2.3.3. Dimerization of MLL chimeras 40
1.2.3.4. Other possible mechanisms 41
1.3. Aims of my projects 43
Chapter 2. Materials and Methods 44
2.1. Antibodies 44
2.2. Stable cell line generation 45
2.3. ES Cell Culture and Differentiation 46
2.4. Lentivirus-based RNAi and Tissue Culture 46
2.5. Flag purification, MudPIT analysis, and size-exclusion chromatography 48
2.6. Immunoprecipitations and kinase assays 49
2.7. Western Blots 50
2.8. Alkaline Phosphatase Characterization 50
2.9. Chromosome Conformation Capture (3C) assay 50
2.10. Electron microscopy 51
2.11. Affymetrix Microarray Analysis 51
2.12. RNAi, RT-PCR, and Total RNA-Seq analysis 51
2.13. ChIP-qPCR and ChIP-Seq Analyses 52
2.14. Track Figures 55
2.15. Histogram and Heatmap Figures 55
2.16. Pol II occupancy Analysis 56

Chapter 3. The Identification of the Super Elongation Complex (SEC) and its role in leukemogenesis 57

3.1. Introduction 57

3.2. Results 58

3.2.1. AFF4 is a shared subunit of several common MLL fusion protein complexes..... 58

3.2.2. AFF4 forms the Super Elongation Complex (SEC) with ELLs and P-TEFb..... 59

3.2.3. SEC is one of the most active P-TEFb-containing complexes..... 62

3.2.4. AFF4 is a central component of SEC..... 65

3.2.5. SEC functions as a transcriptional elongation complex..... 67

3.2.6. AFF4 is required for the proper expression of MLL chimera target genes..... 72

3.2.7. AFF4 is also required for the proper *MYC* gene expression in leukemia cells. 75

3.3. Discussion 76

Chapter 4. P-TEFb within SEC Regulates Rapid Transcriptional Activation in the Presence or Absence of Paused Pol II 80

4.1. Introduction 80

4.2. Results 83

4.2.1. Genome-wide occupancy analyses of SEC components..... 83

4.2.2. SEC is recruited to paused *Hoxa* cluster genes upon RA treatment in ES cells..... 85

4.2.3. SEC is required for the rapid induction of *Hoxa1*..... 88

4.2.4. SEC is required for the induction of other rapidly induced genes. 92

4.2.5. Brd4 is not broadly required for retinoic acid induction of genes..... 95

4.2.6. SEC is required for the rapid induction of immediate early genes (IEG). 96

4.2.7. Dynamic and rapid transcriptional induction requires SEC without the presence of paused Pol II..... 100

4.3. Discussion 102

Chapter 5. The RNA Pol II Elongation Factor Ell3 Marks Enhancers in ES cells and Primes Future Gene Activation 106

5.1. Introduction 106

5.2. Results 109
5.2.1. Ell3 occupies enhancer regions in mouse embryonic stem cells. 109
5.2.2. Ell3 correlates with active, poised, and inactive enhancers. 113
5.2.3. Ell3 is not required for stem cell self-renewal. 114
5.2.4. Ell3 regulates the steady-state expression of a subset of neighboring genes. 115
5.2.5. Ell3 is required for the activation of genes during stem cell differentiation. 118
5.2.6. Ell2 is also required for the activation of some of the Ell3 responsive genes. 119
5.2.7. Ell3 binding at enhancers is required for future gene activation by SEC 120
5.2.8. Ell3 binding at enhancers regulates the Pol II occupancy at promoter-proximal regions of neighboring genes. 121
5.2.9. Ell3-dependent promoter-proximal pausing requires the cohesion complex. 123
5.2.10. Ell3 is essential for stem cell pluripotency and differentiation. 130
5.2.11. Ell3 is present on the chromatin of germ cells. 133
5.3. Discussion 135
5.3.1. A Model for the enhancer-associated Ell3 in the coordinated induction of transcription by SEC 136
5.3.2. Ell3 as a candidate for priming future gene activation 138
5.3.3. A potential role of Ell3 in cancer pathogenesis 140
Chapter 6. Discussion 141
6.1. Different P-TEFb-containing complexes 142
6.1.1. The SEC family of P-TEFb-containing complexes 142
6.1.2. Gene target specificities of the SEC and SEC-like complexes 143
6.2. SEC in rapid gene induction 146
6.2.1. SEC in Pol II-dependent rapid transcriptional activation 146
6.2.2. SEC in Pol II-independent rapid transcriptional activation 147
6.2.3. The potential roles of ELLs in transcriptional initiation control 148
6.3. Mechanisms for SEC recruitment 149
6.3.1. Recruitment of SEC by DNA-specific factors 149

6.3.2. DNA-specific factors independent recruitment of SEC 150

6.4. Target specificities of different elongation factors 153

6.5. Transcription Elongation Checkpoint Control (TECC) and SEC in cancer 153

Chapter 7. Future Work 156

Table of Figures

Figure 1.1. The RNA Pol II transcription cycle. 10

Figure 1.2. Sequence alignment of the ELL family of proteins. 19

Figure 1.3. Distinct forms of P-TEFb-containing complexes. 23

Figure 1.4. Schematic model of key structural domains in wild-type MLL and MLL chimera protein. 30

Figure 1.5. The most common MLL chimeras in acute leukemia 38

Figure 1.6. Proposed mechanisms of MLL chimera-induced leukemogenesis. 42

Figure 3.1. AFF4 is a shared subunit of several of the MLL chimeras and associates with known RNA polymerase II elongation factors 60

Figure 3.2. Schematic representation of the Super Elongation Complex (SEC) 62

Figure 3.3. SEC can phosphorylate Pol II CTD *in vitro*. 63

Figure 3.4. The kinase activities of different P-TEFb-containing complexes 65

Figure 3.5. AFF4 is required for the assembly of SEC-containing ELLs, P-TEFb, and MLL partners. 67

Figure 3.6. Alignment of the AFF1 and AFF4 proteins with *Drosophila* AFF4 (dAFF4). 68

Figure 3.7. The *Drosophila* ortholog of AFF4 colocalizes with ELL and the elongating form of Pol II on *Drosophila* polytene chromosomes. 70

Figure 3.8. AFF4 is required for proper *HSP70* induction 72

Figure 3.9. AFF4/SEC is recruited to MLL chimera target genes in leukemic cells. 74

Figure 3.10. AFF4 is required for the proper expression of MLL chimera target genes in leukemic cells. 75

Figure 3.11. *MYC* expression in leukemia cells is regulated by AFF4. 76

Figure 3.12. Model for SEC in MLL-rearranged leukemia. 78

Figure 4.1. SEC enriched at highly transcribed genes such as the histone loci. 83

Figure 4.2. Global occupancy of the SEC subunits in mouse embryonic stem cells. 84

Figure 4.3. The *Hoxa1* promoter is preloaded with Pol II and recruits SEC after RA treatment in ES cells. 86

Figure 4.4. *Hoxa1*, but not *Hoxb1,* contains the pausing form of RNA Polymerase II in untreated mouse ES cells. 88

Figure 4.5. SEC is required for the rapid induction of the *Hoxa1* gene. 90

Figure 4.6. CDK9 is required for both *Hoxa1* and *Hoxb1* gene activation by RA. 92

Figure 4.7. SEC regulates the rapid induction of retinoic acid signaling. 93

Figure 4.8. SEC is recruited to retinoic acid-induced genes. 94

Figure 4.9. The P-TEFb complex is required for all RA highly induced gene activation. 95

Figure 4.10. Brd4 is broadly present, but not broadly required, for retinoic acid induction of genes. 96

Figure 4.11. SEC is recruited to serum-induced genes. 97

Figure 4.12. SEC is recruited to rapidly induced immediate early genes (IEGs). 99

Figure 4.13. The rapid induction of *Cyp26a1* does not involve pre-loaded Pol II. 101

Figure 4.14. The SEC/P-TEFb complex, but not Brd4/P-TEFb complex, is required for the *Cyp26a1* gene activation by RA. 102

Figure 4.15. Diverse mechanisms for rapid activation of genes during development. 103

Figure 5.1. Ell3 co-localizes with p300 at enhancer regions. 110

Figure 5.2. Ell3 predominantly associates with enhancer regions in mESC. 112

Figure 5.3. Ell3 is found at active, poised, and inactive enhancers. 113

Figure 5.4. Ell3 is not required for the self-renewal of embryonic stem cells. 114

Figure 5.5. Ell3 preferentially regulates the expression of bivalent genes in ES cells. 116

Figure 5.6. Examples of Ell3-associated genes in ES cells. 118

Figure 5.7. Ell3 and Ell2 are both required for the activation of bivalent genes associated with Ell3-occupied enhancers. 119

Figure 5.8. Ell3 binding at enhancers is required for future gene activation by SEC. 120

Figure 5.9. Ell3 regulates Pol II occupancy at promoter-proximal regions of neighboring genes. 122

Figure 5.10. The recruitment of basal transcription factor TFIIB is not affected by Ell3 knockdown. 123

Figure 5.11. Ell3 co-occupies with cohesin and Mediator at enhancers. 124

Figure 5.12. Ell3-mediated promoter-proximal pausing of RNA Pol II requires the cohesin complex. 126

Figure 5.13. Cohesin is not required for the *Ell3* gene expression in mouse ES cells. 127

Figure 5.14. Pol II occupancy analysis after Ell3 knockdown in identified topological domains. 128

Figure 5.15. Ell3 is required for the specific enhancer-promoter interactions in *Hoxa* locus after induction. 129

Figure 5.16. Ell3 is required for the stem cell specification. 131

Figure 5.17. The binding profiles of Ell3, p300, and Pol II on lineage-specific genes. 132

Figure 5.18. Ell3 is essential for the proper neural differentiation of mouse ES cells. 133

Figure 5.19. Ell3 and Pol II localization on the chromatin of germ cells. 134

Figure 5.20. Immunogold labeling of Ell3 in mouse sperm. 135

Figure 5.21. Model for the enhancer-associated elongation factor Ell3 in gene activation 138

Figure 6.1. Distinct classes of genes are regulated by SEC and LEC families from *Drosophila* to human. 143

Figure 6.2. Diverse recruitment mechanisms of SEC in disease. 152

Figure 6.3. SEC in cancer progression. 154

Table of Tables

Table 2.1. List of antibodies used in this thesis. 45

Table 3-1. MudPIT analysis of MLL chimera purifications. 58

Table 3-2. MudPIT analysis of ELL1-3 and AFF4 purifications. 61

Table 6-1. Different P-TEFb-containing complexes in mammals 145

Abbreviations

ADAMTS1	ADAM metallopeptidase with thrombospondin type 1 motif, 1
AdML	adenovirus major late
AF10	ALL1-fused gene from chromosome 10 protein
AF17	ALL1-fused gene from chromosome 17 protein
AF4/ AFF1	ALL1-fused gene from chromosome 4 protein
AF5q31/AFF4	ALL1-fused gene from chromosome 5q31 protein
AF9	ALL1-fused gene from chromosome 9 protein
ALF domain	AF4/LAF4/FMR2 homology domain
ALL	acute lymphoid leukemia
ALYREF	Aly/REF export factor
AML	acute myeloid leukemia
AR	auto regulatory element
ASH2L	ash2 (absent, small, or homeotic)-like
ATF3	activating transcription factor 3
ATP	adenosine 5'-triphoshate
ATPase	adenosine triphosphatase
ATRA	all-trans retinoic acid
Bbx	Bobby sox homolog
BCR	breakpoint cluster region
BRD2	bromodomain-containing 2
BRD3	bromodomain-containing 3
BRD4	bromodomain-containing 4
BRE	TFIIB recognition element
CAK	CDK-activating kinase

CARM1	coactivator-associated arginine methyltransferase 1
CBP	CREB-binding protein
Cbx5	chromobox 5
CDK12	Cyclin Dependent Kinase 12
CDK13	Cyclin Dependent Kinase 13
Cdk6	Cyclin dependent kinase 6
CDK7	Cyclin Dependent Kinase 7
CDK9	Cyclin Dependent Kinase 9
CDKN1A	Cyclin Dependent Kinase inhibitor 1A
Cdx1	caudal-type homeobox 1
ChIP	chromatin immunoprecipitation
ChIP-chip	ChIP followed by microarray hybridization
ChIP-seq	ChIP-sequencing
CPSF160	cleavage and polyadenylation specificity factor 160 kDa subunit
CPSF30	cleavage and polyadenylation specificity factor 30 kDa subunit
CPSF73	cleavage and polyadenylation specificity factor 73 kDa subunit
CSNK1E	casein kinase 1, epsilon
CTCF	CCCTC-binding factor
CTD	C-terminus domain
CTP	cytidine 5'-triphosphate
CUTs	cryptic unstable transcripts
Cyp26a1	cytochrome P450 26A1
DCE	downstream core element
DEAE	diethylaminoethyl cellulose
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease

DOM3Z	dom-3 homolog Z
DOT1L	disruptor of telomeric silencing 1-like
DPE	downstream promoter element
DPY30	DumPY30
DRB	5, 6-dichloro-1-β-D-ribofuranosylbenzimidazole
DSIF	DRB sensitivity-inducing factor
DTT	dithiothreitol
E(z)	enhancer of zeste
EAF1	ELL-associated factor 1
EB	embryoid body
Ebf1	early B cell factor 1
EEN	extra eleven-nineteen leukemia fusion gene
EGF	epidermal growth factor
ELL	Eleven-nineteen Lysine-rich Leukemia gene
ENL	eleven-nineteen leukemia
ENY2	enhancer of yellow 2 homolog
EPS15	epidermal growth factor receptor pathway substrate 15
ESC	embryonic stem cells
FBS	fetal bovine serum
FDR	false discovery rate
Fgf8	fibroblast growth factor 8
FKBP	FK506-binding protein
FMR2/AFF2	fragile X mental retardation 2
FOS	FBJ osteosarcoma oncogene
Foxa2	forkhead box A2
Foxp2	forkhead box P2

GAS7	growth arrest specific 7
GATA1	GATA binding protein 1
Gata4	GATA binding protein 4
Gata6	GATA binding protein 6
Gli1	GLI-Kruppel family member 1
GM-CSF	granulocyte M-CSF
GMP	granulocyte macrophage progenitors
Gro-seq	Global Run-On Sequencing
Gsc	goosecoid homeobox
GST	glutathione S-transferase
GT	guanylyltransferase
H3K4me3	H3 Lysine 4 trimethylation
H4R3	histone H4 arginine 3
HEXIM1	hexamethylene bisacetamide-inducible protein 1
HIV	human immunodeficiency virus
Hmgb3	high mobility group box 3
HOXA10	homeobox A10
HOXA9	homeobox A9
HS2	hypersensitivity site 2
HSC	hematopoietic stem cells
HSP26	heat shock protein 26
Hsp70	heat shock protein 70
ICE1	Interact with C-terminus ELL 1
ICE1	Interact with C-terminus ELL 2
IDT	Integrated DNA Technologies
IEG	immediate early genes

Igh	immunoglobulin heavy chain
INR	initiator element
Irx3	iroquois homeobox 3
Kcnc1	potassium voltage-gated channel subfamily C member 1
KLH	keyhole limpet hemocyanin
Lacz	β-galactosidase
LAF4/AFF3	lymphoid nuclear protein related to AF4
LARP7	La-related protein
LEC	Little Elongation Complex
LEDGF	lens epithelium-derived growth factor
Lefty1	left-right determination factor 1
LIF	leukemia inhibitory factor
LSCs	leukemia stem cells
LTR	long terminal repeats
MAT1	menage a trois-1
MED26	mediator momplex subunit 26
MEF	mouse embryonic fibroblast
MEIS1	Meis homeobox 1
MEN1	multiple endocrine neoplasia 1
MePCE	methyl phosphate capping enzyme
MLL	mixed lineage leukemia
MT	guanine-7-methyltransferase
MTE	motif ten element
MudPIT	multidimensional protein identification technology
Myb	myeloblastosis oncogene
NEAA	nonessential amino acids

NEB	New England Biolabs
NELF	negative elongation factor
NFE2	nuclear factor, erythroid-derived 2
Ni-NTA	nickel-nitriloacetic acid
nt	nucleotide
Oct4	octamer-binding protein 4
Olig3	oligodendrocyte transcription factor 3
PAF1	RNA polymerase II-associated factor 1 homolog
PCF11	pre-mRNA cleavage complex II
PHD	Plant Homeo Domain
PIC	pre-initiation complex
Pou5f1	POU class 5 homeobox 1
PP1	protein phosphatase 1
PRMT1	protein arginine methyltransferase 1
P-TEFb	Positive Transcription Elongation Factor b
RAP30	RNA polymerase II-associating protein 30
RAP74	RNA polymerase II-associating protein 74
RAR	RA receptor
RARE	retinoic acid response element
RbBP5	retinoblastoma binding protein 5
Rere	arginine glutamic acid dipeptide (RE) repeats
RNA Pol I	RNA polymerase I
RNA Pol II	RNA polymerase II
RNA Pol III	RNA polymerase III
RNA	ribonucleic acid
RNGTT	RNA guanylyltransferase and 5'-phosphatase

RNP	ribonucleoprotein
rNTP	ribonucleotide tri-phosphate
RNU6	U6 small nuclear RNA
Rpb1	DNA-directed RNA polymerase II subunit1
RPKM	Reads Per Kilobase of transcript per Million
RRM	RNA recognition motif
rRNA	ribosomal RNA
Runx1	runt-related transcription factor 1
SAM68	Src-associated in mitosis of 68 kDa
SARs	scaffold attachment regions
SCAF4	SR-related CTD-associated factor 4
SCAF8	SR-related CTD-associated factor 8
SEC	Super Elongation Complex
SEC-L2	SEC-like 2
SEC-L3	SEC-like 3
SEPT6	septin 6
SETD1A	SET domain-containing 1A
SETD1B	SET domain-containing 1B
SLBP	stem-loop binding protein
Smc1a	structural maintenance of chromosomes 1A
Smc3	structural maintenance of chromosomes 3
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
Sox2	SRY box-containing gene 2
SRSF1	serine/arginine-rich splicing factor 1
SRSF2	serine/arginine-rich splicing factor 2

Su(Tpl)	suppressor of triplo lethal
T	brachyury
TBP	TATA-box binding protein
Topo II	topoisomerase II
tRNA	transfer RNA
Trx	trithorax
TSS	transcription start site
UBF	upstream binding transcription factor, RNA polymerase I
UV	ultraviolet
WDR5	WD repeat domain 5
XPB	xeroderma pigmentosum, complementation group B
XPD	xeroderma pigmentosum complementary group D
XRN2	5'-3' exoribonuclease 2
YEATS	Ynl107, ENL, AF9, and TFIIF small subunits

Chapter 1. Introduction

1.1. Transcriptional elongation control by RNA polymerase II

The control of gene expression underlies almost all cellular events such as cell signaling and communication in development and pathogenesis. In eukaryotes, the regulation of transcription is a multi-faceted and highly regulated process involving a large number of regulatory factors. Transcription starts with the assembly of the transcriptional initiation machinery on the promoter of a gene, followed by transcriptional elongation and termination. The initiation process is relatively well characterized and plays pivotal roles in transcriptional regulation. Recent evidence, especially genome-wide studies, also demonstrated a general and essential role of the elongation stage in regulating the proper expression of developmentally controlled genes. This introduction mainly covers the transcription of protein-coding genes by RNA polymerase II, introduces elongation factors including the Positive Transcription Elongation Factor b (P-TEFb) and the Eleven-nineteen Lysine-rich Leukemia (ELL) family of proteins, and discusses the role of ELL in mixed-lineage leukemia (MLL) chimera-induced leukemogenesis.

1.1.1. RNA polymerases in mammals

1.1.1.1. The identification of three mammalian RNA polymerases

DNA-dependent RNA polymerase, which can specifically and efficiently synthesize RNA in the presence of ribonucleotide tri-phosphate (rNTP) by using DNA as a template, was first discovered by Weiss (Weiss and Gladstone, 1959). In 1959, Weiss' group observed the incorporation of [α-^{32}P] CTP (Cytidine 5'-triphosphate) into ribonucleic acid (RNA) in the rat liver nuclei, and the incorporation efficiency was greatly affected upon the addition of Deoxyribonuclease (DNase) (Weiss and Gladstone, 1959).

One year later, Stevens and Hurwitz also reported that *Escherichia coli* extracts can support the incorporation of ribonucleotides into RNA (Hurwitz et al., 1960; Stevens, 1960). By 1961, the first DNA-dependent RNA polymerase was successfully isolated from *Micrococcus lysodeikticus* (Weiss and Nakamoto, 1961).

In prokaryotes, there is only one RNA polymerase, which consists of four core subunits ($\alpha 2\beta\beta'$) and one auxiliary factor (δ) and is responsible for synthesizing all types of RNA (Yura and Ishihama, 1979). In eukaryotic cells, there are three different nuclear DNA-dependent RNA polymerases that are responsible for transcribing distinct classes of genes. By using Diethylaminoethyl cellulose (DEAE)-Sephadex chromatography, Roeder and Rutter first separated the three mammalian RNA polymerases under different concentrations of ammonium sulfate and named them as RNA polymerase I, II, and III (Roeder and Rutter, 1969, 1970a, b). RNA polymerase I (RNA Pol I) is responsible for the transcription of ribosomal RNA (*rRNA*); RNA polymerase II (RNA Pol II) transcribes all the protein-coding genes, microRNA, and most of the small nuclear RNA (*snRNA*); RNA polymerase III (RNA Pol III) synthesizes transfer RNA (*tRNA*), 5s rRNA, and some small RNAs, like small nuclear RNA 7SK and U6 small nuclear RNA (RNU6).

The RNA polymerases are complex and multi-subunit enzymes each consisting of 12-17 polypeptides with some subunits common to all of the three enzymes (Cramer et al., 2008). Substrate specificities for each RNA polymerase are partially reflected by these complex subunit compositions. Early studies showed that the purified RNA polymerases lack the intrinsic capabilities to specifically initiate transcription at the core promoter of a specific class of genes (Sentenac, 1985). A variety of transcription factors, including general transcription factors and regulatory factors, were later identified to cooperate with the enzymes to ensure accurate transcription initiation in these class-specific genes. For

example, in the reconstituted, *in vitro* transcription system, transcription initiation factors IA TIF-IA, -IB, -IC, and UBF (upstream binding transcription factor, RNA polymerase I) are required for RNA Pol I to specifically initiate the transcription at the rDNA gene promoters (Haltiner et al., 1986; Hanada et al., 1996; Learned et al., 1986; Smale and Tjian, 1985); and general transcription factors TFIID, TFIIB, TFIIA, TFIIE, TFIIF, and TFIIH for the initiation of RNA Pol II on protein-coding genes (Buratowski et al., 1989; Conaway et al., 1987; Conaway and Conaway, 1989a; Conaway and Conaway, 1989b; Davison et al., 1983; Dynlacht et al., 1991; Inostroza et al., 1991; Maldonado et al., 1990; Matsui et al., 1980; Ohkuma et al., 1990; Samuels et al., 1982; Tanese et al., 1991; Zhou et al., 1991).

1.1.1.2. The structure of RNA polymerase II

RNA polymerase II has been purified to near homogeneity in many eukaryotes, including yeast, fly, and human. In mammals, RNA Pol II consists of 12 evolutionarily conserved subunits, including 5 'core' subunits (conserved in all cellular organisms), 5 common subunits (common in all three Pol I, II, and III), and two other subunits, Rpb4/7 (DNA-directed RNA polymerase II subunit 4/7) (Cramer et al., 2008). The necessity of each subunit to the cell's viability has been well characterized in the model organism *Saccharomyces cerevisiae*. DNA-directed RNA polymerase II subunit 1 (Rpb1), subunit 2 (Rpb2), subunit 3 (Rpb3), subunit 5 (Rpb5), subunit 6 (Rpb6), subunit 8 (Rpb8), subunit 10 (Rpb10), and subunit 11 (Rpb11) are essential for the viability of yeast cells (Young, 1991). Although the Rpb4/7 and Rpb9 deletion strains are viable, these cells usually grow slowly and are sensitive to extreme temperature conditions (Woychik and Young, 1990).

Over the past forty years, much effort has been put into deciphering the mechanism of RNA Pol II transcription, including the composition and assembly of the Pol II

holoenzyme, the enzyme-substrate interaction, and the regulation of accurate transcription initiation. In 2001, Kornberg and colleagues reported high-resolution atomic structures of the free and elongating forms of yeast RNA polymerase II, which is highly relevant to mammalian Pol II (Cramer et al., 2001; Gnatt et al., 2001). The 10-subunit core enzyme can be divided into four mobile modules, including core module (Rpb1, 2, 3, 10, 11, and 12), shelf module (Rpb1 cleft, Rpb1 foot, Rpb5, and Rpb6), clamp (Rpb1 clamp core and clamp head, Rpb2 clamp), and the jaw-lobe module (Rpb1 jaw, Rpb9 jaw, and Rpb2 lobe). The positively charged active center cleft is formed by the largest subunits Rpb1 and Rpb2 within the core module. A bridging helix extending from the Rpb1 subunit spans over the cleft, lining a "pore" in the active center. The entrance of the cleft is located in the area between the "lower" (Rpb5) and the "upper" (Rpb9) jaw. The clamp, which regulates the opening and closing of the cleft, controls the entering of the DNA template into the active center. There are two Magnesium ions found within the active center cleft named "metal A" and "metal B". Metal A binds to the highly conserved "aspartate loop" of Rpb1 in the active center, and metal B is 5.8 Å away from metal A, separated by the Rpb1 aspartate loop.

In the structure of the transcribing RNA Pol II complex (Gnatt et al., 2001), the downstream DNA template enters into the Pol II complex from the entrance, travels along the bottom of the clamp, and then passes over the bridging helix in the active center cleft. Around 4 base pairs of the unwound DNA template form the transcription preceding the 3'-hydroxyl terminus of the synthesizing RNA chain, followed by the 9 base pair of the DNA-RNA hybrid formed in the active center cleft. Once the RNA transcript reaches 10 residues, it releases from the DNA-RNA hybrid and enters into the groove located at the base of the Rpb1 CTD, where it exits the polymerase complex. In the active center, the bridging helix directly interacts with the coding base of the DNA template, which helps

position the nascent base pair. Furthermore, the extensive contacts between the clamp and the DNA-RNA hybrid induce dramatic conformation changes at the base of the clamp, possibly leading to the swinging of the clamp over the cleft and then trapping the DNA template and the nascent transcript.

The catalytic mechanism of transcription by RNA polymerase II is not very clear. Thomas A. Steitz and colleagues proposed the two-metal ion mechanism (Steitz, 1998), which is supported by the structures of bacterial and yeast elongating Pol II complexes. In this model, in the catalytic active center, metal A and B bind to the phosphate group of the 3' end of the RNA and NTP (nucleotide triphosphate) triphosphate moiety, respectively (Cramer et al., 2008; Cramer et al., 2001; Westover et al., 2004). The delivery of NTP into the insertion site is then induced by the folding of the trigger loop next to the bridging helix (Westover et al., 2004).

1.1.1.3. RNA polymerase II large subunit Carboxyl Terminal Domain

Unlike RNA Pol I and III, the largest subunit of RNA Pol II has a unique C-Terminal Domain (CTD) consisting of the conserved and tandemly repeated heptapeptides: Threonine (Tyr1)- Serine (Ser2)- Proline (Pro3)- Threonine (Thr4)- Serine (Ser5)- Proline (Pro6)- and Serine (Ser7) (Corden, 1990). The consensus sequence of the CTD is evolutionarily conserved with variable repeats in different species, such as 26 repeats in yeast, 45 in fly, and 52 in mammals (Conaway and Conaway, 1993; Young, 1991). *In vivo*, there are at least two distinct forms of RNA Pol II depending on the phosphorylation status of the CTD: II0 for the hyper-phosphorylated form and IIA for the hypo-phosphorylated form (Dahmus, 1995). Early ultraviolet (UV) cross-linking studies from Dahmus' group suggested that the II0 form is predominantly associated with the

elongating RNA Pol II, while the initiating Pol II is the hypo-phosphorylated IIA form (Dahmus, 1995). Although the CTD is essential for cell viability in yeast, fly, and mammals, it is not widely required for the *in vitro* transcription initiation assay using many different promoters like the adenovirus major late (AdML) promoter (Buratowski and Sharp, 1990; Kim and Dahmus, 1989; Phatnani and Greenleaf, 2006; Young, 1991).

Ser2, Thr4, Ser5, and Ser7 within the CTD consensus sequence can all be phosphorylated both *in vivo* and *in vitro*. The Cyclin-Dependent Kinase 7 (CDK7) within the TFIIH complex can phosphorylate the CTD Ser5, whereas Ser2 is phosphorylated by Cyclin-Dependent Kinase 9 (CDK9) within the positive transcription elongation factor, P-TEFb, complex and the recently identified Cyclin-Dependent Kinase 12 (CDK12) complex with Cyclin K (Bartkowiak et al., 2010; Blazek et al., 2011; Feaver et al., 1994; Hengartner et al., 1998; Liu and Kipreos, 2000; Marshall et al., 1996; Phatnani and Greenleaf, 2006; Roy et al., 1994; Serizawa et al., 1995; Shiekhattar et al., 1995; Sun et al., 1998). CDK9 is also responsible for the phosphorylation of Thr4 (Hsin et al., 2011). Genome-wide distribution analyses using phosphorylation-specific antibodies indicate that all the Ser2, Ser5, and Ser7 phosphorylation forms of Pol II enrich at the active protein-coding genes, with Ser5 and Ser7 peaking at the 5' end and Ser2 phosphorylation at the 3' end (Kim et al., 2010; Mayer et al., 2010; Tietjen et al., 2010).

In the past 15 years, a large number of studies have demonstrated that the CTD phosphorylation might serve as a platform for the recruitment or stable binding of regulatory factors to chromatin, coupling nuclear events with transcription (Orphanides and Reinberg, 2002; Reed, 2003) (Discussed later). The CTD deletion mutant of RNA Pol II greatly inhibits the RNA processing *in vivo*, including splicing, 3'-end processing, and termination (McCracken et al., 1997). By affinity chromatography, McCracken and

colleagues further demonstrated that the cleavage/polyadenylation factor, CPSF, and the cleavage stimulation factor, CSTF, can directly interact with the glutathione S-transferase (GST) tagged CTD (GST-CTD) (McCracken et al., 1997). A number of CTD-interacting proteins have been identified so far, mostly by yeast-two hybridization and GST-CTD affinity chromatography. Many of these proteins are involved in various RNA-processing or nuclear pathways, including the 5'-end mRNA capping (RNA guanylyltransferase and 5'-phosphatase (RNGTT), mRNA cleavage/polyadenylation (pre-mRNA cleavage complex 2 protein (PCF11), cleavage, and polyadenylation specificity factor 160 kDa subunit (CPSF160), cleavage and polyadenylation specificity factor 73 kDa subunit (CPSF73), cleavage and polyadenylation specificity factor 30 kDa subunit (CPSF30), protein phosphatase 1 (PP1), Symplekin), transcription termination (5'-3' exoribonuclease 2 (XRN2), Senataxin, SR-related CTD-associated factor 8/4 (SCAF8/4), and dom-3 homolog Z (DOM3Z)), mRNA export (enhancer of yellow 2 homolog (ENY2), and Aly/REF export factor (ALYREF)), splicing (serine/arginine-rich splicing factor 1 (SRSF1), and serine/arginine-rich splicing factor 2 (SRSF2)), DNA damage and repair (casein kinase 1, epsilon (CSNK1E)), and histone modifications (histone lysine methyltransferase Set1, RNA polymerase II-associated factor 1 homolog (PAF1) and Set2) (Finkel et al., 2010; Kuehner et al., 2011; Morris and Greenleaf, 2000; Pascual-Garcia et al., 2008; Phatnani et al., 2004).

Recent studies from Reinberg's group demonstrated that the arginine R1810 on the Pol II CTD non-consensus repeats is methylated by coactivator-associated arginine methyltransferase 1 (CARM1), regulating the expression of snRNA and snoRNA (small nucleolar RNA) genes (Sims et al., 2011). The methylation of R1810 by CARM1 is inhibited by Ser2/5 phosphorylation *in vitro*, leading to the hypothesis that this residue is methylated before transcription initiation. However, how R1810 methylation effects the

sn/snoRNA gene expression or transcription initiation on sn/snoRNA genes still remains unclear.

Besides the phosphorylation and methylation, CTD can also be glycosylated on the serine and threonine residues (Kelly et al., 1993); glycosylation and phosphorylation are mutually exclusive. However, no *in vivo* functional evidence for this modification has been reported to date. The unique YSPTSPS consensus sequence suggests that CTD could be the substrate of the prolyl isomerase, Pin1 (Ess1 in yeast), which can catalyze the *cis/trans*-isomerization of the prolyl-peptide bond after the phosphorylation of the proceeding serine/threonine (Fanghanel and Fischer, 2004; Lu and Zhou, 2007). Studies from the Vincent (Albert et al., 1999) and Hanes (Wu et al., 2000) groups indicated that indeed both mammalian Pin1 and yeast Ess1 can bind to the phosphorylated Pol II CTD. The binding of Pin1 to CTD directly inhibits the CTD phosphatase FCP1-mediated dephosphorylation of RNA Pol II by directly inhibiting the FCP1 activity (Palancade et al., 2004; Xu et al., 2003). Like CTD glycosylation, however, it is still unclear whether the isomerization of phosphorylated Serine/Threonine-Proline can also behave as a platform to recruit other factors to regulate transcription, RNA processing, or other nuclear events.

1.1.2. Transcription by RNA polymerase II

1.1.2.1. The assembly of the pre-initiation complex

Transcription by RNA Pol II can be divided into several steps, starting from the assembly of the pre-initiation complex (PIC), followed by the activation of initiation, promoter clearance, transcription elongation, and termination (Figure 1.1). The assembly of the pre-initiation complex involves RNA Pol II, basal transcription factors, and the core promoter DNA (Conaway and Conaway, 1993; Roeder, 2005; Sikorski and Buratowski,

2009). Unlike the bacterial RNA polymerase, eukaryotic RNA Pol II cannot accurately initiate the transcription from natural templates *in vitro* (Axel et al., 1973; Gilmour and Paul, 1973; Parker et al., 1978). Early studies from Roeder and colleagues demonstrated that accessory factors from crude cellular extracts are required for accurate transcription initiation by RNA Pol II from the major late viral promoter *in vitro* (Matsui et al., 1980; Weil et al., 1979). Many accessory factors (collectively named as basal/general transcription factors) were later biochemically purified by the chromatography fractionation, including TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (Conaway and Conaway, 1993; Roeder, 1996; Thomas and Chiang, 2006; Zawel and Reinberg, 1993). For the core promoter DNA, at least 6 different elements have been identified so far, which are required for the proper and accurate assembly of the pre-initiation complex, including the TATA box, downstream promoter element (DPE), initiator element (INR), TFIIB recognition element (BRE), motif ten element (MTE), and the downstream core element (DCE) (Smale and Kadonaga, 2003).

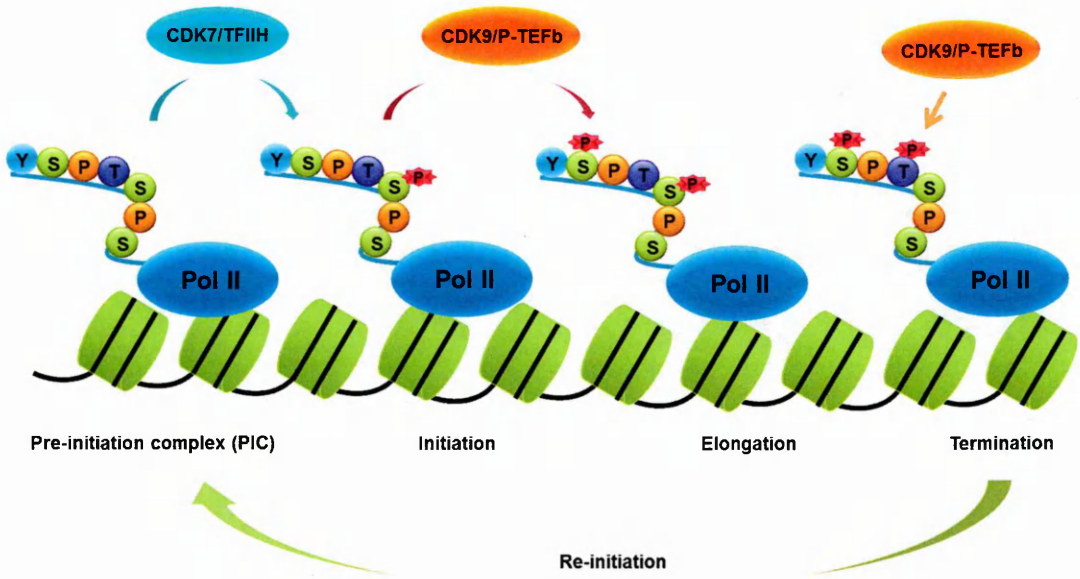


Figure 1.1. The RNA Pol II transcription cycle.
Transcription by RNA Pol II can be divided into four steps: assembly of the pre-initiation complex (PIC), initiation, elongation, and termination. In general, transcription starts with the assembly of hypophosphorylated RNA Pol II with the general transcription factors forming the PIC. The initiating form of the Pol II complex is marked by CDK7 within TFIIH-mediated CTD Ser5 phosphorylation, whose level is reduced in the elongating form of Pol II. During the transition from initiation into elongation, Ser2 on Pol II CTD is further phosphorylated by CDK9 within P-TEFb, leading to the release of Pol II into productive elongation. Unlike Ser5 phosphorylation, Ser2 phosphorylation is increased in the gene body and peaks at the end of the transcript unit. Threonine 4, involved in the 3'-end processing of histone genes, can also be phosphorylated by CDK9. Pol II falls from the template after the termination of transcription and can be re-incorporated into the PIC.

Earlier studies from Roeder's and Sharp's groups indicated that the assembly of the initiation complex might follow a multistage or sequential binding process *in vitro* (Fire et al., 1984; Hawley and Roeder, 1985; Samuels and Sharp, 1986). This sequential assembly model was further defined by Buratowski and colleagues from the Sharp laboratory at the AdML promoter through a native gel electrophoresis DNA binding assay (Buratowski et al., 1989). The saddle-shaped TATA-box binding protein (TBP) within the TFIID complex first binds to the TATA box located at the minor groove of the core promoter and then bends the promoter DNA around 90 degrees, providing the platform for the assembly of PIC. The TFIID-DNA complex is further stabilized by the entry of TFIIB directly

contacting both the TBP and the core promoter BRE sequence with its C-terminus. The N-terminus of the TFIIB consisting of a zinc ribbon motif can interact with the RNA Pol II subunits Rpb1 and Rpb2, and also the RAP30 subunit of the TFIIF complex, facilitating the recruitment of RNA Pol II and TFIIF to the TFIIB-TFIID-promoter ternary complex. TFIIF contains two subunits, the RNA polymerase II-associating proteins 30 (RAP30) and 74 (RAP74), which can interact with different subunits of the RNA Pol II complex, facilitating the formation and stabilization of the Pol II-TFIIB-TFIID-promoter complex. TFIIF is also necessary for the recruitment of TFIIE and TFIIH, probably by directly interacting with TFIIE. Once the TFIIE and TFIIH are recruited to the TFIIF-Pol II-TFIIB-TFIID-promoter complex, the assembly of PIC is complete.

1.1.2.2. Early elongation

Upon the addition of rNTP, RNA Pol II synthesizes the first phosphodiester bond. Early short transcripts are often unstable and easily released from the promoter, resulting in RNA Pol II re-initiating in a process called "abortive initiation" (Dvir, 2002; Hsu, 2002). It has been suggested that in this stage the Pol II complex will experience three transitions called "promoter clearance" including the 4 nucleotide (nt), 10 nt, and 15 nt stages (Cramer, 2004). Adenosine 5'-triphoshate (ATP), TFIIE, and TFIIH are continuously required at these stages (Kugel and Goodrich, 1998). This early transcribing complex becomes stable once the nascent transcript reaches 15 nucleotides.

The TFIIH complex, containing a core module (TFB5, P34, P44, P52, P62, and the XPB (xeroderma pigmentosum, complementation group B) and XPD (xeroderma pigmentosum complementary group D) helicases) and a CDK-activating kinase (CAK) module (CDK7, Cyclin H and MAT1 (menage a trois-1)), was initially purified from rat

liver extracts by the Conaway group (Conaway and Conaway, 1989b; Ranish et al., 2004). So far, three distinct enzymatic activities have been observed for the TFIIH complex: DNA-dependent adenosine triphosphatase (ATPase), ATP-dependent helicase, and CTD kinase activities (Thomas and Chiang, 2006). It has been demonstrated that the DNA-dependent ATPase activity of the TFIIH complex is required for opening of promoter and synthesizing the first phosphodiester bond (Holstege et al., 1996). Studies from Reinberg's laboratory indicated that TFIIH is also required for efficient promoter escape as, in the absence of TFIIH, the transcription complex often pauses at the promoter-proximal region (Kumar et al., 1998).

1.1.2.3. Promoter-proximal pausing

During the transition from the pre-initiation to early elongation stage, RNA Pol II transcribes 20~40 nucleotides and then pauses at the promoter-proximal region (Fuda et al., 2009). At this stage, RNA Pol II is phosphorylated on CTD Ser5 by CDK7. The studies from the Svejstrup laboratory showed that the phosphorylated CTD Ser5 can disrupt Pol II-mediator interaction, resulting in the release of the Pol II from the mediator complex and possibly further promoting the early transcribing complex escaping from the core promoter (Max et al., 2007). The phosphorylated CTD Ser5 is also involved in the recruitment of the 5'-end RNA capping enzyme, which is required for the stability of the nascent transcript (Ho et al., 1998; Komarnitsky et al., 2000; Schroeder et al., 2000).

Early studies of the bacterial RNA polymerase suggested that destabilization of the DNA-RNA hybrid can cause backtracking of RNA polymerase along the template, which is referred to as the back and forth movement of the polymerase along DNA and RNA (Komissarova and Kashlev, 1997; Nudler et al., 1997). This backtracking can further

cause transcriptional pausing or arrest (Cramer, 2004). In the arrested stage, the 3'-end of the RNA transcript extrudes and is not properly aligned with the active site of the polymerase (Gilmour, 2009). The transcript cleavage factor, TFIIS, was initially identified to promote RNA synthesis by RNA Pol II (Sekimizu et al., 1976). Further *in vitro* evidence and the Pol II-TFIIS structure demonstrated that TFIIS prevents the transcriptional arrest through promoting both cleavage of the nascent transcript by RNA Pol II and also the realignment of the transcript 3'-end with the active site (Izban and Luse, 1992b; Kettenberger et al., 2003).

The *Drosophila* heat shock protein 70 (*Hsp70*), mammalian v-myc myelocytomatosis viral oncogene homolog (*MYC*), FBJ osteosarcoma oncogene (*FOS*), and the human immunodeficiency virus (*HIV*) genes were the first examples to demonstrate paused Pol II at their promoter-proximal regions around 20-40 nucleotides downstream from the transcription start site (TSS) or with transcriptional elongation blocks at the first exon of the gene (Collart et al., 1991; Gilmour, 2009; Krumm et al., 1992; Laspia et al., 1989). It is now widely accepted that besides the pre-initiation step, the elongation step, especially the releasing of the polymerase from the paused or arrested state, also plays a pivotal role in regulating gene transcription (Levine, 2011; Smith et al., 2011a). Genome-wide RNA Pol II occupancy analyses from *Drosophila* embryos and S2 cells reveal that ~ 10% of the *Drosophila* genome, mostly developmentally regulated genes, are poised or repressed, to be activated at the later developmental stages, largely expanding the function of promoter-proximal pausing in transcription control (Muse et al., 2007; Zeitlinger et al., 2007). Other related studies regarding the Pol II distributions in murine embryonic stem cells showed that the vast majority of the protein-coding genes contain RNA Pol II at their promoters with peaks at the promoter-proximal regions (Guenther et al., 2007). These results have led to the argument that the promoter-proximal

pausing is a general mechanism in controlling the transcription of all genes, including house-keeping genes, stress-responsive, and developmentally regulated genes with differences in the frequency of the RNA Pol II releasing from promoter regions (Nechaev and Adelman, 2008).

The establishment of RNA Pol II pausing

Factors and regulatory elements play central roles in setting up paused Pol II at promoter-proximal regions. The importance of regulatory elements in regulating the paused Pol II was first shown in early studies analyzing the promoter-proximal architecture of the *Drosophila Hsp70* gene (Lee et al., 1992). The deletion or mutation of the GAGA element upstream from the *Hsp70* promoter significantly reduced the level of paused Pol II, possibly through affecting the recruitment of the nucleosome remodeling factor, the Nurf complex, also suggesting the critical function of regulatory element binding factors in the regulation of paused Pol II (Gilmour, 2009; Lee et al., 1992; Tsukiyama and Wu, 1995). However, later studies on the Heat shock protein 26 (*HSP26*) promoter indicated that the GAGA factor is also required for the recruitment of basal machinery TFIID, arguing that the GAGA factor might actually affect the transcription initiation or both (Gilmour, 2009; Lu et al., 1993; O'Brien et al., 1995; Sandaltzopoulos et al., 1995; Wall et al., 1995). Other independent studies on the *immunoglobulin Ig Kappa* gene indicated the importance of regulatory elements in setting up paused Pol II at promoter-proximal regions by deleting the intron and C region of the *Ig Kappa* gene, which contain enhancer elements for this gene (Raschke et al., 1999).

Reduction of long RNA species, but the accumulation of the short capped transcripts after the treatment of the ATP analog 5, 6-dichloro-1-β-D-

ribofuranosylbenzimidazole (DRB), indicates that this small molecule is able to suppress productive elongation without affecting initiation (Chodosh et al., 1989). By using DRB to inhibit transcriptional elongation, Handa and colleagues identified the factors that negatively regulate the elongation stage: the DRB sensitivity-inducing factor (DSIF) and the negative elongation factor (NELF) (Wada et al., 1998; Yamaguchi et al., 1999). Both DSIF and NELF are required for the induction of the promoter-proximal pausing of RNA Pol II *in vitro*, while neither DSIF nor NELF alone have this function (Renner et al., 2001; Yamaguchi et al., 2002; Yamaguchi et al., 1999). DSIF mediates the interaction between NELF and RNA Pol II. NELF is a four-subunit complex and the C-terminus of its E subunit contains a RNA recognition motif (RRM), which can bind to RNA (Yamaguchi et al., 1999). Mutations of this RRM affect the ability of NELF to repress transcriptional elongation without disrupting the formation of the NELF complex and the interaction of NELF-DSIF-RNA Pol II. These findings suggested that NELF cooperates with DSIF to induce the pausing state of RNA Pol II at the promoter-proximal region through binding to RNA Pol II and the nascent transcript (Gilmour, 2009; Yamaguchi et al., 2002; Yamaguchi et al., 1999).

This model is supported by the *in vivo,* protein-DNA cross-linking assays showing that DSIF and NELF co-occupy the promoter-proximal regions of many paused genes, for example, the *Hsp70* gene, the immediate early genes, and the provirus *HIV* (Aida et al., 2006; Andrulis et al., 2000; Ping and Rana, 2001; Wu et al., 2003). Upon induction, NELF rapidly dissociates from or remains at the promoter-proximal region of *Hsp70* or *FOS*, while DSIF together with Pol II travels into the gene body, supporting the additional positive roles of DSIF in elongation (Wu et al., 2005; Yamada et al., 2006). Consistent with these observations, genome-wide localization of DSIF and NELF in mouse embryonic stem cells demonstrated that these two factors peak at the promoter-proximal

region with RNA Pol II at both actively transcribed and non-productive genes, while DSIF also travels with RNA Pol II into the gene body in actively transcribed genes (Rahl et al., 2010). Functional studies of NELF and DSIF demonstrated that depletion of the NELF subunit E reduces the paused Pol II levels at the *Hsp70* gene promoter and this effect was later observed in 115 genes out of the 200 paused *Drosophila* genes by Chromatin immunoprecipitation (ChIP) followed by microarray hybridization (ChIP-chip), and that knockdown of Spt5 in the DSIF complex leads to the increase of RNA Pol II levels in the gene body (Muse et al., 2007; Rahl et al., 2010; Wu et al., 2003).

Factors involved in reducing transient pausing

Several factors involved in reducing transient pausing have been biochemically purified based on their abilities to promote the catalytic activities of RNA Pol II on naked DNA templates *in vitro*, including TFIIF, Elongin, and ELL (Aso et al., 1995; Bradsher et al., 1993a; Bradsher et al., 1993b; Flores et al., 1989; Izban and Luse, 1992a; Price et al., 1989; Shilatifard et al., 1996; Tan et al., 1994). The early studies performed in bacteria suggested that the destabilization of the DNA-RNA hybrid causes the backtracking of RNA polymerase (Komissarova and Kashlev, 1997; Nudler et al., 1997), which will induce the transcriptional pausing or arrest. Therefore, it is possible that the proper alignment of the 3' end of the RNA transcript with the active site of RNA Pol II would prevent transcriptional pausing or arrest. Indeed, *in vitro* evidence demonstrated that TFIIF, Elongin, and ELL can increase the catalytic rate of transcription elongation by Pol II by suppressing transient pausing (Shilatifard et al., 2003). In the following section, the current understanding of the *in vivo* functional evidence for the involvement of TFIIF, Elongin A and ELL in transcriptional elongation control will be discussed.

TFIIF

TFIIF is essential for the assembly of the pre-initiation complex (Price et al., 1989; Thomas and Chiang, 2006). Although both *Drosophila* and mammalian TFIIF have the ability to promote the elongation rate of RNA Pol II, the TFIIF mutant, which lacks the elongation activity, shows the defect in reducing the frequency of abortive transcription initiation by Pol II (Yan et al., 1999). In addition, genome-wide ChIP-sequencing (ChIP-seq) analysis of TFIIF in yeast revealed that TFIIF almost exclusively occupies the promoter, but not the coding region, supporting the major role of TFIIF in modulating initiation and early elongation events (Rhee and Pugh, 2012). The function of TFIIF in elongation was recently further substantiated by the finding that the Pol II tightly associated, repressing factor, Gdown1, competes with TFIIF for binding with Pol II, and thus, causes the pausing of RNA Pol II at promoter-proximal regions (Cheng et al., 2012; Jishage et al., 2012).

Elongin A

The Elongin complex, consisting of the active module, Elongin A, and the E3 ubiquitin ligase, Elongin B/C, is responsible for the polyubiquitination and further degradation of the RNA Pol II larger subunit I in response to DNA damage signals in both yeast and human (Harreman et al., 2009; Ribar et al., 2007; Yasukawa et al., 2008). It is possible that Elongin is needed to remove the polymerase from where it pauses under stress conditions. In *Drosophila*, it has also been demonstrated that Elongin A relocalizes to the heat shock puff sites upon stress and is required for the proper induction of the *Hsp70* gene (Gerber et al., 2005a).

ELLs

The ELL gene was originally identified in patients suffering from acute myeloid leukemia (AML) as one of the frequent fusion partners of the *MLL* gene (Shilatifard et al.,

2003; Smith et al., 2011b; Thirman et al., 1994). The characterization of ELL as a transcription elongation factor has led to the proposal that the misregulation of the elongation stage of transcription could play an important role in MLL chimera-mediated leukemogenesis (Shilatifard et al., 1996). In mammals, there are three ELL family members, ELL1-3, all showing activities of stimulating transcription elongation *in vitro* (Figure 1.2). Northern blot analyses indicated that both *ELL1* and *ELL2* are ubiquitously expressed in all kinds of tissues, while *ELL3* shows testis specific expression (Miller et al., 2000; Shilatifard et al., 2003).

```
ELL    ----MAALKEARSYGLSCGRVSDGSRVSVFHVKLTDSALKAFESYRAHQDSVSLRPSIRFEGSQ 60
ELL2   MAAGGAAGLREEQRYGLACGRLGQDN-ITVLHVKLTETAIRALETYQSHKNLIPFRPSIQFQGLQ 64
ELL3   ----MEGTQEALS-GKMRLLFTPAARTSLLMLRLNEAALRALQECQQQQ----VRPVIAFQGHR 55
                    Elongation stimulation domain
ELL    GHISIPQP--DCPEEVRAFSFYLSNIGRDSPQGSFDCIQQYVSSYGDVHLDCLGSIQDKVTVCAT 123
ELL2   GLMKIPKN--DPFNEVQNFNFYLSNVGRDNPQGSFDCIQQTLSSSGASQLNCLGFIQDKITVCAT 127
ELL3   GYLRFPGPGWSCL-----FSFIVSQCGQEGTNGGLDLVYQRLGRSGPNCLHCLGSLRERLTIWAA 115

ELL    DDSYQKARQSMAQAEEETRSRSAIVIKAGGRYMGKKVQFRKPAPGAADAVPSRKRATPINLASAI 188
ELL2   NDSYQMTRERMTQAEEESRNRSTKVIKPGGPYVGKRVQIRKAPQAISDTVPERKRSTPMNPANTI 192
ELL3   ----------------------------------------------MDTIP-----APL------ 123

ELL    RKSSGSGASSVVQRPFRDRVLHLLALRPYRKAELLLRLQKDGLTQADKDTLDSLLQQVASVNPKD 253
ELL2   RKM--HSGNSVSQRPYRDRVIHLLALKAYKKPELLARLQKDGVNQKDKNSLGAILQQVANLNPKD 255
ELL3   ------------------------------------LAQEHLTEGTRESES------------- 138

ELL    GTCTLQDCMYKSLQKDWPGYSEGDRQLLKRMLMRKLCQPQNATT--DSSPPREHGRSASPSQKRP 316
ELL2   LSYTLKDYVFKELQRDWPGYSETDRQTLDLVLSRKLNPSQNASTSRSESPLCSSKDAASSPQKRP 320
ELL3   ----------------WQDTGD-----------------------------------EPEGHPQLAP 154

ELL    --TDFIDPLASKKPRISHFTQRAQPTLNGKLGAPNGHET--LLPAPGPTPSDTLS---SSHLPPR 374
ELL2   LDSDFIDPLMNKKARISHLTNRVPPTLNGYLNPTSEKSCAGLLPPPAAAAIPTLSPLPSTHLPVS 385
ELL3   --DEVSDPLASHHEQ--------------------------SLPGSSSEPMAQWEMRNHTYLPSR 191

ELL    LEPPRTHDPLADVSNDLGHSTQDY-------------KHQEATPAPAPHLGLPLLTDFPQAEQPT 426
ELL2   NPPQTVNSNSNSPSTPEGLGTQDLPVDSFSQNGSIFEDQQEKYTSRTCLETLPPSSALLKCPKPM 450
ELL3   EPDQSLLSPASQKRL----------------------DKKRSAPITTEEPEEKRLRALPLASSPL 234

ELL    SSSH--THSRPKKKSKKHKDKER---------PPEERPPAPQPDAPTAPALPPDAPGLNGACDN- 479
ELL2   EEEHPVSHKKSKKKSKKHKEKDQIKKLDIETMEEKEEDLQREETAKLSNASPNPNEGVKEGCTAS 515
ELL3   QGLA-------NQDSQEGEDWGQ----------DEDEEGDEDGDSRLEQSLSAPSAS-------- 274

ELL    -EPTSSSETPDYLLKYPAISSSEQRQSYKNDFNAEYSEYRSLHARIEQITRRFTQLDAQLRQLSQ 543
ELL2   MEPSSALELPDYLIKYIAIVSYEQRQNYKDDFNAEYDEYRALHARMETVARRFIKLDAQRKRLSP 580
ELL3   -ESPSPEEVPDYLLQYRAIHSTEQQQAYEQDFETDYAEYRILHARVGAASQRFTELGAEIKRLQR 338
                    Occludin homology domain
ELL    GSDEYETTRGQILQEYRKIKKTNTNYSCEKRRCEYLHRKLAHIKRLIAEYQQRQLQAWP 602
ELL2   GSKEYQNVHEEVLQEYQKIKQSSPNYHEEKYRCEYLHNKLAHIKRLIGEF--------- 630
ELL3   GTPEHKVLEDKIVQEYKKFRKRYPSYREEKHRCEYLHQKLSHIKGLILEFEEKNRGS-- 395
```

Figure 1.2. Sequence alignment of the ELL family of proteins.

In mammals, there are three ELL family proteins, ELL1, ELL2, and ELL3, with a conserved N-terminal elongation stimulation domain and a C-terminal occludin homology domain. All of the three ELL proteins have stimulating activities on transcription elongation by RNA Pol II *in vitro*. *ELL1* and *ELL2* are expressed ubiquitously; whereas *ELL3* shows testis-specific expression by Northern blot analysis.

The ELL family contains a conserved N-terminal domain and a C-terminal occludin homology domain. The N-terminus of ELL can interact with RNA Pol II and ELL-associated factor 1/2 (EAF1/2) and is essential for stimulating transcriptional elongation(Kong et al., 2005). The binding of EAF1/2 to ELL greatly enhances the elongation activity of ELL *in vitro* (Kong et al., 2005). EAF1 and ELL co-localize at the Cajal bodies, and this colocalization is transcription-dependent as the inhibition of transcription by α-amanitin or DRB disperses their localizations into the whole nucleus

(Polak et al., 2003). It has been suggested that Cajal bodies are the sites for snRNA genes transcription and RNA processing (Gall et al., 1999), indicating a potential role of ELL/EAF on snRNA transcription or RNA processing. The N-terminus of ELL can also interact with tumor protein p53 (Shinobu et al., 1999), possibly mediating ELL overexpression-induced programmed cell death (Johnstone et al., 2001). ELL inhibits both p53-dependent transcriptional activation and repression through sequence-dependent and independent manners, respectively (Shinobu et al., 1999). In turn, p53 can also inhibit ELL's stimulatory activity on Pol II elongation.

In the MLL-ELL translocation (further discussed in section 1.2), the extreme C-terminal (497-621aa), but not the N-terminal (1-373aa), region of ELL is necessary and sufficient to immortalize the primary murine hematopoietic progenitor cells in a colony formation assay (DiMartino et al., 2000). Interestingly, engineered Mll-Eaf1 can also induce AML (Luo et al., 2001), suggesting an important role of EAF1 in MLL-ELL-mediated leukemogenesis. Although the p53 interaction region of ELL is not necessary for MLL-ELL-mediated leukemogenesis, MLL-ELL inhibits p53-mediated apoptosis and also *cyclin-dependent kinase inhibitor 1A* activation (*CDKN1A*) (Wiederschain et al., 2003). It is likely that the inhibition of p53-mediated apoptosis by MLL-ELL could in turn promote its induction of leukemia.

In *Drosophila*, there is only one ELL homologous protein, dEll, encoded by the *Suppressor of triplo lethal* (*Su(Tpl)*) gene (Shilatifard et al., 2003). Mutations of dEll are recessively lethal and the heteroallelic combinations of these mutations cause embryonic segmentation defects (Eissenberg et al., 2002). Subsequent truncation analyses of dEll demonstrated that the N-terminus of dEll is sufficient for its localization to transcriptionally active puff sites, while the overexpression of its C-terminus can rescue the

recessive lethality phenotype of dEll mutations (Gerber et al., 2005b). The first *in vivo* evidence that Ell might serve as an elongation factor came from *Drosophila* studies. Polytene staining indicated that dEll is associated with phosphorylated Pol II at many actively transcribed loci (Eissenberg et al., 2002). Like P-TEFb, dEll is also rapidly relocalized to the heat shock puff sites upon stress and required for the proper induction of the *Hsp70* gene in the *Drosophila* salivary gland. This result provides *in vivo* evidence for the possible role of ELL in reducing paused Pol II (Smith et al., 2008).

In addition, RNAi studies in *Drosophila* indicated that dEll is required for the global Ser2 phosphorylation on Pol II CTD, which is visualized by polytene staining (Smith et al., 2008). Interestingly, reduction of Cdk9, which is required for the Ser2 phosphorylation in flies, also affects the chromatin localization of dEll (Eissenberg et al., 2007). These results suggested that CTD-phosphorylated Pol II is required for the recruitment of dEll to chromatin, while the localization of dEll to chromatin can further enhance Pol II phosphorylation by Cdk9.

Besides *Drosophila* and human, ELL is also found in other metazoan species, including cow, cat, dog, pig, sheep, marmoset, chicken, and fish (Thirman et al., 1994). Recent studies from the Conaway group indicated that the lower eukaryote *Schizosaccharomyces pombe* expresses ELL (spELL) and EAF (spEAF) homologs (Banks et al., 2007). Like their mammalian homologs, spELL and spEAF can also stimulate the transcriptional elongation by RNA Pol II. However, the deletion mutants of both spELL and spEAF are viable. Like many transcription elongation factors, the spELL mutant is sensitive to the 6-azauracil.

1.1.2.4. P-TEFb and productive elongation

Following the hypothesis that the early elongation complex might be both negatively and positively regulated, Price and colleagues identified the positive transcription elongation factor, P-TEFb, which can promote the rate of long transcript synthesis by RNA Pol II (Peterlin and Price, 2006). P-TEFb is a two-subunit complex, composed of CDK9 and cyclin T1/2. *In vitro* elongation assays demonstrated that P-TEFb can eliminate the inhibitory effect of DSIF and NELF on elongation, possibly by phosphorylating Pol II CTD on the Ser2 and E subunit of NELF (Fujinaga et al., 2004; Kim and Sharp, 2001; Renner et al., 2001; Yamada et al., 2006). DSIF also contains the consensus repeats Gly-Ser-Arg/Gln-Thr-Pro in its C-terminus, which is similar to the Pol II CTD consensus repeat and can be phosphorylated by P-TEFb (Yamada et al., 2006). The phosphorylation of DSIF is required for the induction of the immediate early gene *FOS* by the epidermal growth factor (EGF) and also for the travel of Pol II into the gene body. Therefore, it has been proposed that P-TEFb can phosphorylate Pol II CTD, NELF, and DSIF resulting in the dissociation of NELF from the promoter-proximally paused Pol II and the further release of the paused Pol II into productive elongation (Zhou et al., 2012).

The co-existence of both the active and inactive forms of the P-TEFb complexes *in vivo* allows the transition of P-TEFb from one state to another to reach the functional equilibrium, as required by the cellular real time needs (Zhou et al., 2012) (Figure 1.3). The majority of the P-TEFb complex is stored within the inactive 7SK snRNP complex. The 7SK/P-TEFb complex contains hexamethylene bisacetamide-inducible protein 1/2 (HEXIM1/2), La-related protein, LARP7, and the 7SK methyl phosphate-capping enzyme, MePCE. HEXIM1/2, the RNA-binding protein, can bind to 7SK RNA, and the resulting 7SK-HEXIM1/2 complex can further interact with the phosphorylated T-loop region of CDK9, keeping the P-TEFb complex in an inactive state (Li et al., 2005). 7SK snRNP (small nuclear ribonucleoprotein) is diffused into the nucleus, providing the P-TEFb complex to any active chromatin loci (Zhou et al., 2012). However, studies on the *HIV* promoter indicated that the P-TEFb complex can be recruited to the promoter within the inactive 7SK snRNP complex, leading to the inhibition of elongation (D'Orso and Frankel, 2010). This result also suggested that the transition from an inactive to active state of the P-TEFb complex could happen at the promoters of genes.

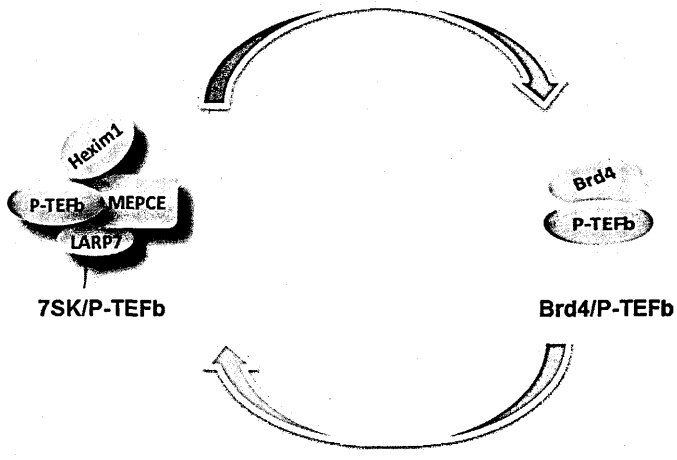


Figure 1.3. Distinct forms of P-TEFb-containing complexes. P-TEFb is a two-subunit complex, consisting of cyclin-dependent kinase 9 (CDK9) and cyclin T1/2. P-TEFb can phosphorylate Serine 2 on Pol II CTD, NELF, and SPT5, resulting in the release of paused Pol II from the promoter-proximal region. *In vivo,* the majority of the P-TEFb complex is stored within the inactive 7SK snRNP complex. The co-existence of both active (BRD4/P-TEFb) and inactive (7SK/P-TEFb) forms of the P-TEFb complexes allows the transition of P-TEFb from one state to another to reach functional equilibrium.

Bromodomain-containing 4 (BRD4) was found to form an active complex with P-TEFb, which can phosphorylate the Pol II CTD at the Serine 2 residues *in vitro* (Jang et al., 2005; Yang et al., 2005). BRD4 is recruited to the paused *HIV-1 LTR* (*Long terminal repeats*) promoter. However, BRD4 is only required for the basal expression, but not the Tat-mediated transactivation of *LTR*. Overexpression of BRD4 results in the suppression of the Tat-mediated transactivation as BRD4 competes with Tat for binding with P-TEFb (Bisgrove et al., 2007). BRD4 is required for *MYC* gene expression, and that the depletion of BRD4 reduces the localization of CDK9 to the chromatin (Yang et al., 2008). It is possible that the main function of P-TEFb within the BRD4 complex is involved in regulating the basal expression of genes. However, it still remains elusive which P-TEFb-containing complex is responsible for the release of paused Pol II during rapid gene activation.

1.1.2.5. Transcriptional elongation-coupled RNA processing

Ser2 phosphorylated Pol II CTD is one of the critical marks for productive elongation and itself can also serve as a recruitment platform for some of the elongation/RNA processing-related complexes, which are able to facilitate the smooth travelling of the elongating Pol II throughout the transcript unit by creating an open chromatin environment or promoting RNA processing and/or maturation. There is much evidence that has demonstrated that RNA processing is tightly coupled with the Pol II transcribing process, starting from the addition of a 7-methylguanosine cap at the 5' end, splicing of introns, and lastly, the cleavage and polyadenylation of nascent transcript at the 3'-end.

RNA capping

The 7-methyl G5'ppp5'N cap of the nascent Pol II transcript occurs soon after the nascent RNA reaches a length of 25-30 nt, emerging from the exit channel of Pol II (Shuman, 1997). The capping process requires the sequential enzymatic actions catalyzed by three different enzymes RNA tri-phosphatase, guanylyltransferase (GT), and the guanine-7-methyltransferase (MT). Both GT and MT enzymes can specifically bind to the CTD phosphorylated form of Pol II (Bentley, 2002). *In vivo* studies from yeast demonstrated that Kin28, which can phosphorylate Ser5 at Pol II CTD, is required for the recruitment of all three capping enzymes (Komarnitsky et al., 2000; Schroeder et al., 2000). Further *in vitro* studies with synthetic CTD peptides also indicated that the Ser5, but not Ser2, phosphorylated Pol II CTD can stimulate the activity of GT by inducing the allosteric change of GT (Ho and Shuman, 1999). The RNA tri-phosphatase Pct1 and GT can also bind to the SPT5 subunit of the DSIF complex, which in turn stimulates the capping (Wen and Shatkin, 1999). The link between nascent transcript capping and early elongation suggests a 'checkpoint' mechanism exists to ensure the proper capping before Pol II enters into the productive elongation (Bentley, 2002).

Pre-mRNA splicing

A number of studies have demonstrated that the pre-mRNA is at least partially spliced during the transcriptional process as a significant amount of spliceosome assembles at the spliced sites of transcribing genes (Carrillo Oesterreich et al., 2011). The first example of co-transcriptional splicing was observed in *Drosophila;* it is when nascent *chorion* gene transcripts are shortened during transcription (Beyer and Osheim, 1988). Also, the electron microscopy imaging of the spread chromatin showed the assembly of ribonucleoprotein particles at spliced sites of nascent transcripts (Osheim et al., 1985). Recent nascent RNA-sequencing studies in yeast provide genome-wide evidence that the

majority of yeast transcribing genes are co-transcriptionally spliced (Carrillo Oesterreich et al., 2010). However, it still remains unclear whether introns are completely removed during transcription.

It has been suggested that Pol II elongation rates can influence the alternative splicing. Pol II mutations or drug inhibition, which slows down the elongation rate of Pol II, results in the distinct alternative exon inclusion (Carrillo Oesterreich et al., 2011). Supporting this observation, the hyperphosphorylated Pol II CTD can interact with splicing factors, like SR protein SF2/ASF and SC35, leading to the recruitment of key splicing factors to the elongation complex, and possibly, the further assembly of a spliceosome complex (Zhou et al., 2012). Interestingly, recent experiments by increasing the active pool of P-TEFb complexes demonstrated that the reduction of the inactive P-TEFb complex component LARP7 or MePCE promotes the inclusion of the alternative exon (Barboric et al., 2009). It is possible that the increased active pool of the P-TEFb complexes enhances the levels of Ser2 phosphorylation and also the Pol II elongation rate, which in turn elevates the SR protein-mediated assembly of the splicing complex at exons.

3' end processing

Poly (A)-dependent and Sen1-dependent termination pathways are two, well-characterized pathways for Pol II-transcribed protein-coding and non-coding genes, respectively (Kim et al., 2006; Shandilya and Roberts, 2012). In eukaryotes, most of the protein-coding genes contain poly (A) signal (5'- AAUAAA-3') and GU rich sequences after the cleavage site. Multiple lines of evidence have indicated that Ser2 phosphorylated Pol II CTD can serve as a platform to recruit termination factors, including CPSF and CstF, promoting the transcription-coupled 3'-end processing (Ahn et al., 2004). CPSF is recruited through its interaction with the body of Pol II of the transcribing complex, while

the further recruitment of CstF is dependent on Ser2-phosphorylated Pol II CTD (Nag et al., 2007; Shell et al., 2007). *In vivo* studies from Martinson's laboratory demonstrated that Pol II reduces the elongation rate after transcribing through the poly (A) signal and thus pauses after the poly (A) site (Park et al., 2004). Therefore, it is possible that the binding of CPSF to the AAUAAA signal on the transcripts can induce the pausing of Pol II, which is released by the subsequent binding of CstF with the GU-rich sequences and the further recruitment of other 3'-end processing factors like the Rat1-Rai1-Rtt103 complex (Kuehner et al., 2011). Paused Pol II per se does not necessarily cause the transcriptional termination. The "Torpedo" model has been proposed. In this model, the interaction of Rt103 with Pol II CTD recruits the 5'-3' exoribonuclease, Rat1, which cleavages nascent RNA from the uncapped end generated by other 3'-end endoribonuclease towards the RNA exit channel on Pol II, chasing down Pol II, and thus inducing transcriptional termination (Kuehner et al., 2011). Recently, CDK9-mediated Thr 4 phosphorylation on Pol II CTD was also found to be required for the 3'-end processing of histone genes by recruiting the CPSF-100 and stem loop binding protein (SLBP) (Hsin et al., 2011).

The Sen1-dependent termination pathway is required for the termination of snRNA, small nucleolar RNA (snoRNA) and Cryptic Unstable Transcripts (CUTs) in *S. cerevisiae* (Arigo et al., 2006; Finkel et al., 2010; Kim et al., 2006; Thiebaut et al., 2006). However, the depletion of Sen1 does not seem to affect the 3'-end processing of the snRNA genes in mammals (Suraweera et al., 2009). Instead, the integrator complex was found to mediate this process (Baillat et al., 2005). The association of the integrator complex with snRNA genes requires the Ser7 phosphorylated Pol II CTD (Egloff et al., 2007). Disruption of Ser7 on Pol II CTD affects the 3'-end processing and also termination of snRNA genes, but not protein-coding genes, for example, the housekeeping gene, *Actin*.

Recently ELL2 has also been shown to regulate the polyadenylation site selection of the *immunoglobulin heavy chain* (*Igh*) gene in plasma cells through directing the localization of the polyadenylation factor, CstF-64, to the *Igh* locus (Martincic et al., 2009). Depletion of ELL2 by shRNA reduces the ratio of secretory (shorter): membrane-specific (longer) *Igh* transcripts. Further, reporter assays indicated that the reduction of ELL2 might affect the exon skipping, resulting in the use of the promoter-proximal poly (A) site. It is possible that the depletion of ELL2 slows down the elongation rate of RNA Pol II, and therefore, provides enough time for the assembly of the splicing machinery at the weaker alternative splice site.

1.2.MLL rearrangement-mediated leukemogenesis

The pathogenesis of cancer is a multi-step process initiating from chromosomal translocations and genetic mutations, which generally provide a proliferation advantage to the cancer-initiating cells, but impair their ability to undergo terminal differentiation and apoptosis. Early studies demonstrated that chromosomal translocations, resulting in the activation of proto-oncogenes or more often the novel fusion genes with acquired functions, can be a causative factor in hematopoietic malignancies (Rowley, 2001; Showe and Croce, 1987). For example, the t(8; 14) (q24; q32) translocation, which predominates in Burkitt lymphoma, results in the activation of oncogene *MYC* as it is juxtaposed to a highly active immunoglobulin heavy-chain locus in B cells (Showe and Croce, 1987; Taub et al., 1982). Translocations involving the *MLL* gene are found frequently in *de novo* and secondary leukemias, representing 80~90% of infant acute lymphoid leukemia (ALL) and ~40% of infant acute myeloid leukemia (Meyer et al., 2009). Almost all MLL translocations produce functional chimera proteins, consisting of the amino terminus of MLL fused to the carboxyl terminus of fusion partners. Mice transplanted with MLL-

translocated stem/progenitor cells or *in vivo* knock-in with MLL translocations develop leukemia after 3~6 months of latency. However, the mechanisms underlying how MLL translocations lead to leukemogenesis are as diverse as its fusion partners.

1.2.1. *MLL* encodes a histone H3 lysine 4 methyltransferase

1.2.1.1. *MLL* is a homolog of the *Drosophila trithorax* gene

The Trithorax group and polycomb group of proteins are functionally and genetically connected, controlling the activation and repression of genes during differentiation (Ringrose and Paro, 2004). Mutations in *trithorax* (*Trx*) cause the homeotic transformation defects in *Drosophila*, which phenocopy the defects observed in the mutants of homeotic genes *antennapedia* and *ultrabithorax* (Breen and Harte, 1993; Castelli-Gair and Garcia-Bellido, 1990; Eissenberg and Shilatifard, 2010; Shearn, 1989; Wedeen et al., 1986). MLL, encoding a 3,969 amino acid (aa) protein, is structurally homologous to the *Drosophila trithorax*. Knockout studies demonstrated that the deletion of *MLL* or its SET domain causes the axis skeleton defects in mice through regulating the expression of the *Hox* cluster genes (Terranova et al., 2006; Yagi et al., 1998; Yu et al., 1995). However, unlike the MLL deletion, SET domain deletion mice are viable, suggesting additional important functions of MLL in development besides the roles of the SET domain in regulating *Hox* gene expression. MLL knockout mice also showed the defects in hematopoietic systems, possibly affecting the development of hematopoietic stem cells in a *Hox* gene-dependent manner (Yagi et al., 1998).

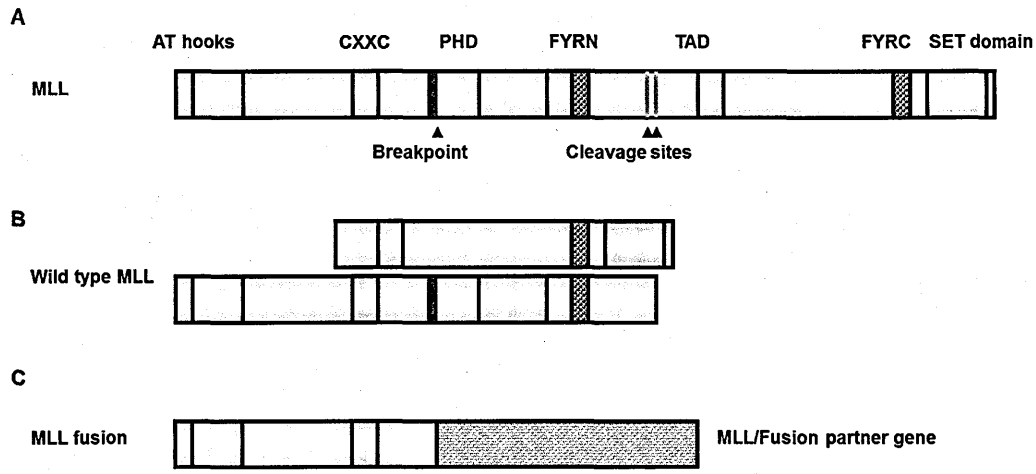


Figure 1.4. Schematic model of key structural domains in wild-type MLL and MLL chimera protein. (A) MLL is a 3,969 amino acid protein which contains several identified domain structures, including the methyltransferase SET domain, several PHD fingers, DMT homology CXXC domain, and AT hooks. AT hooks and CXXC domains can bind to DNA and are responsible for the recruitment of MLL or MLL chimera proteins. (B) Wild-type MLL can be cleaved and forms a heterodimers of MLL N-terminus and MLL C-terminus through FY-rich FYRN and FYRC domains. Recently, the PHD fingers were also shown to be involved in the dimerization of MLL protein. (C) MLL chimeras are formed by the in-frame fusion of the N-terminus of MLL to the C-terminus of its translocation partners. The AT hook and CXXC domains of MLL are retained in all chimeras. MLL rearrangements mostly occur in the breakpoint cluster region (BCR) between exons 8 and 11. To date, more than 60 fusion partners have been identified with low to no sequence similarities.

MLL contains sequence-independent DNA binding motifs at the N-terminus, including three AT-hook domains binding to the AT-rich sequences and CXXC motifs binding to CG-rich sequences (Birke et al., 2002; Zeleznik-Le et al., 1994) (Figure 1.4). The first 50 aa of MLL, which can interact with MENIN encoded by the *multiple endocrine neoplasia 1* (*MEN1*) gene, together with AT-hook domains and CXXC motifs are required for the recruitment of MLL to the chromatin (Yokoyama and Cleary, 2008). It has also been reported that CXXC motifs of MLL can interact with the PAF1 complex, resulting in the recruitment of MLL (Milne et al., 2010; Muntean et al., 2010). Besides these two motifs, the N-terminus of MLL also contains several PHD fingers (Plant Homeo Domain). It has been reported that the third PHD finger can bind to the di-and trimethylated histone H3 lysine 4, possibly regulated by cyclophilin CyP33 (Wang et al.,

2010b). The recognition of the MLL PHD finger on the histone H3 Lysine 4 trimethylation (H3K4me3) mark is essential for its transcriptional activities towards *homeobox A9* (*HOXA9*), *A10* (*HOXA10*), and the *Meis homeobox 1* (*MEIS1*) genes, but not its localization to chromatin (Chang et al., 2010). The C-terminus of MLL contains a methyltransferase enzymatic domain, the SET domain, which was named after the *Drosophila* SET domains containing proteins *Su(var)*3-9, Enhancer of zeste (*E(z)*), and *Trx* (Milne et al., 2002; Nakamura et al., 2002).

Wild-type MLL is cleaved by taspase 1 into MLL-NT and MLL-CT fragments that form a heterodimer endogenously through the interaction of FYRN and FYRC domains, respectively (Hsieh et al., 2003; Yokoyama et al., 2002) (Figure 1.4). Recent studies indicated that the PHD fingers are also required for the dimerization of MLL (Yokoyama et al., 2011). *In vivo*, the free MLL-NT and -CT fragments are subjected to different degradation pathways (Yokoyama et al., 2011). The MLL-CT is required for the stability of the MLL-NT fragment. However, it remains unclear if the proteolysis of mammalian MLL is necessary for some of its biological functions. In *Drosophila*, the mutant of *Trithorax* that contains a deletion spanning the cleavage sites by taspase 1 affects the maturation of the *Trx* protein and also the *antennapedia*, but not the *bithorax*, gene expression (Mazo et al., 1990). Mice with the deletion mutant of taspase 1 show the homeotic transformation phenotype (Takeda et al., 2006). Also, the deletion of taspase 1 affects the progression of cell cycle, possibly through regulating cell cycle genes that are targets of MLL.

1.2.1.2. MLL forms a COMPASS-like complex

Yeast Set1, sharing the homologous SET domain with mammalian *MLL* and *Drosophila Trx*, is the first identified histone H3 Lysine 4 (H3K4) methyltransferase (Miller et al., 2001; Shilatifard, 2012). Set1 within its complex COMPASS (complex of proteins associated with Set1) is capable of mono-, di-, and trimethylating the H3K4 both *in vivo* and *in vitro* (Krogan et al., 2002; Miller et al., 2001). Unlike yeast, there are at least six functionally nonredundant H3K4 methyltransferases found in human, including SET domain-containing 1A/1B, SETD1A/B, and MLL1-4. Biochemistry characterization of complexes together with functional studies indicated that SETD1A and SETD1B complexes, which are responsible for the bulk levels of H3K4 trimethylation, are the direct homologs of yeast COMPASS with the similar composition (Lee and Skalnik, 2005; Lee et al., 2007; Wu et al., 2008). MLL1-4 form COMPASS-like complexes sharing core subunits with human COMPASS (Cho et al., 2007; Goo et al., 2003; Hughes et al., 2004; Nakamura et al., 2002). Retinoblastoma-binding protein 5 (RbBP5), WD repeat domain 5 (WDR5), ash2 (absent, small, or homeotic)-like (ASH2L), and DumPY30 (DPY30) are the four SET domain-interacting core subunits, with each of them having conserved roles in modulating the H3K4 methyltransferase activities from yeast to human (Dou et al., 2006; Steward et al., 2006).

Besides these core subunits, COMPASS and COMPASS-like also contain unique components, which are now believed to be required for the diverse recruitment of these different complexes to chromatin. MENIN and lens epithelium-derived growth factor (LEDGF) interact with the N-terminus of MLL1/2, forming a ternary complex respectively, which is required for the recruitment of the complex to the chromatin and also transcription of its target genes (Wang et al., 2009; Yokoyama and Cleary, 2008).

Recent studies demonstrated that both MENIN and LEDGF are essential for the MLL chimera-induced leukemogenesis (Agarwal et al., 1999; Yokoyama and Cleary, 2008). However, the expression of MLL2 chimeras, which also contain the interaction domain with MENIN and LEDGF, in hematopoietic stem/progenitor cells, cannot induce leukemia, suggesting the unique function of MLL in leukemogenesis (Bach et al., 2009).

In both *Drosophila* and mammals, COMPASS and COMPASS-like complexes show different activities on the substrate H3K4 *in vivo* and regulate the expression of distinct subset of genes (Shilatifard, 2012). COMPASS-like MLL3/4 complexes mediate the activation of hormone receptor responsive genes (Lee et al., 2009). MLL1 regulates the expression of *Hox* genes and also cell cycle regulatory genes (Liu et al., 2009; Mohan et al., 2010b). Our recent studies further demonstrated that Mll-mediated H3K4 trimethylation is only required for the expression of a small subset of genes in MEFs, including many *Hox* genes and genes involved in the Wnt signaling pathway (Wang et al., 2009). Interestingly, it has also been reported that the Wnt signaling pathway is required for the self-renewal of leukemia stem cells of AML induced by a MLL chimera (Wang et al., 2010a; Yeung et al., 2010).

1.2.2. Characteristics of MLL-translocated leukemia

1.2.2.1. 11q23 rearrangements

The *MLL* gene was initially identified and cloned from the breakpoint of chromosome 11q23 translocations by several groups in 1992 (Djabali et al., 1992; Gu et al., 1992; Tkachuk et al., 1992). The majority of MLL rearrangements occur within an 8.3 Kb region spanning from exons 8 to 13, named 'Breakpoint Cluster Region (BCR)'(Rowley, 1998). Early studies indicated that the chemotherapy treatment with

epipodophyllotoxins, an inhibitor of DNA topoisomerase II (Topo II), often induces 11q23 translocations in cancer patients (Hunger et al., 1993; Pui et al., 1989; Super et al., 1993), suggesting a possible role of an inappropriate repair of DNA double-strand break in chromosome translocations (Zhang and Rowley, 2006). BCR contains the Topo II cleavage sites, DNase I hypersensitive sites and also the scaffold attachment regions (SARs), all of which might contribute to the chromosomal rearrangements at 11q23 (Broeker et al., 1996). Currently, the "transcription factories" model has also been suggested to explain the translocations of MLL with a diverse range of fusion partners (Rowley, 2008).

1.2.2.2. MLL chimera lacks the conserved enzymatic SET domain

The translocations of MLL result in the loss of the methyltransferase SET domain, the dimerization FY-rich FYRN and FYRC domains, and the PHD fingers (Figure 1-4). Although MLL translocations lose the methyltransferase activity, the germline copy of MLL is normally recruited to its target genes, including *HOXA9* and *MEIS1*, and is also required for the induction of leukemia by MLL chimeras (Thiel et al., 2010). Inclusion of the PHD fingers prevents MLL chimera-mediated leukemogenesis (Muntean et al., 2008), possibly by inhibiting the dimerization of MLL chimeras. However, there is no evidence showing the existence of the dimerization between MLL chimera and germline MLL.

The MENIN/LEDGF interaction domain, AT-hooks, and CXXC domain at the N-terminus of MLL are retained in MLL translocations (Birke et al., 2002; Slany et al., 1998; Zeleznik-Le et al., 1994). The MLL fragment of the translocations is required for the recruitment of the chimera complex to chromatin, while the fusion fragment of the

translocations can recruit elongation factors and chromatin modifiers, or induce the dimerization of MLL chimeras, resulting in leukemogenesis.

1.2.2.3. Gene expression signature

Early expression array analyses indicated that the expression profiles of MLL-rearranged leukemia cells are significantly distinct from ALL or AML cells with germline MLL (Armstrong et al., 2002; Rozovskaia et al., 2001). About 200 highly up-regulated genes were identified in this study, including both lymphoid and myeloid specific genes, suggesting the multi-lineage gene expression pattern of MLL-rearranged leukemia cells (Armstrong et al., 2002). Several homeobox-containing genes, such as *HOXA9* and *MEIS1*, are consistently up-regulated in most of the MLL-rearranged leukemia cells (Armstrong et al., 2002; Rozovskaia et al., 2001). Co-expression of *Hoxa9* and *Meis1*, but not the *Hoxa9* or *Meis1* alone, in hematopoietic stem cells (HSC) can induce acute myeloid leukemia (Kroon et al., 1998). The exact functions of HOXA9 and MEIS1 in MLL chimera-induced leukemogenesis are not very clear. It is likely that the overexpression of a cohort of *Hoxa* cluster genes and *Meis* genes are important for the proliferation of leukemia cells or the phenotypes of leukemia, but not the development of MLL chimera induced leukemia (Ayton and Cleary, 2003; Kumar et al., 2004; So et al., 2003b; Zeisig et al., 2004).

Besides homeobox-containing genes, MLL chimera might also activate some developmental regulators involved in the self-renewal of leukemia stem cells (LSCs) (Cleary, 2009). MLL-ENL can induce leukemia from committed myeloid progenitors, granulocyte macrophage progenitors (GMP), suggesting that the expression of MLL-ENL can somehow reactivate some key regulators involved in the stemness and self-renewal of

HSC (Cozzio et al., 2003; So et al., 2003a). The leukemic induction of MLL-ENL on GMP is not directly mediated through the activation of *Hoxa9* and *Meis1* genes, as the co-expression of these two proteins was shown not to be able to induce leukemia from GMP, but HSC (Kroon et al., 1998). ChIP-seq studies demonstrated that the chimera protein MLL-AF4 occupies a subset of genes, which are differentially expressed in HSC, in MLL-AF4-rearranged leukemia cells (Guenther et al., 2008).

1.2.2.4. The transcriptome of MLL LSC is more akin to ESC than HSC

MLL LSCs were first reported in mouse AML induced by MLL-AF9 (Somervaille and Cleary, 2006). The transcription profile of MLL LSC is more related to embryonic stem cells (ESC) when compared with adult stem cells, HSC (Krivtsov et al., 2006; Somervaille et al., 2009). Unlike ESCs, *Nanog*, *POU class 5 homeobox 1* (*Pou5f1*), and the *SRY-box-containing gene 2* (*Sox2*) are not required for the self-renewal of LSCs. Instead, *Hoxa* and *Meis* are required for the proliferation of LSCs. Interestingly, the co-expression of the *myeloblastosis oncogene* (*Myb*), *high mobility group box 3* (*Hmgb3*), and *chromobox 5* (*Cbx5*), which are expressed in both ESC and LSC, can induce immortalization of myeloid progenitors in a *Hoxa9*/*Meis1*-independent manner. These data indicated that at least two parallel or cooperative programs contribute to MLL chimera-mediated leukemogenesis. It is possible that MLL chimera might induce myeloid progenitor cells into a pre-leukemia stem cell (preLSC) stage through the *Myb*, *Hmgb3,* and *Cbx5* activation program, while the further transition into and maintaining the LSC stage requires the *Hoxa* and *Meis* program.

1.2.3. Diverse mechanisms for MLL chimera-mediated gene expression

To date, more than 60 different MLL fusion partners have been identified with little or no obvious sequence similarities (Meyer et al., 2009). Some of them can be simply classified into different groups based on their different cellular localizations, ranging from nuclear to cytosol and membrane-associated proteins (Krivtsov and Armstrong, 2007; Meyer et al., 2009). Translocations with nuclear proteins are found in the majority of MLL-rearranged leukemias (Figure 1.5). For example, fusion partners like the ALL1-fused gene from chromosome 4 protein (AF4/ AFF1), ALL1-fused gene from chromosome 9 protein (AF9), eleven-nineteen leukemia (ENL), ALL1-fused gene from chromosome 10 protein (AF10), ELL, and the ALL1-fused gene from chromosome 17 protein (AF17) are all nuclear proteins and are the most frequent translocation partners with different frequencies in different types of leukemia (Mohan et al., 2010b). It is not clear whether MLL translocations involving proteins within the same group will induce leukemia in a similar manner or whether translocations with nuclear proteins will involve different mechanisms from fusions with cytosol or membrane-associated proteins during the induction of leukemogenesis. Currently, several models have been proposed of how MLL chimera oncoproteins can lead to leukemogenesis or how these oncoproteins activate the gene expression program involved in leukemogenesis.

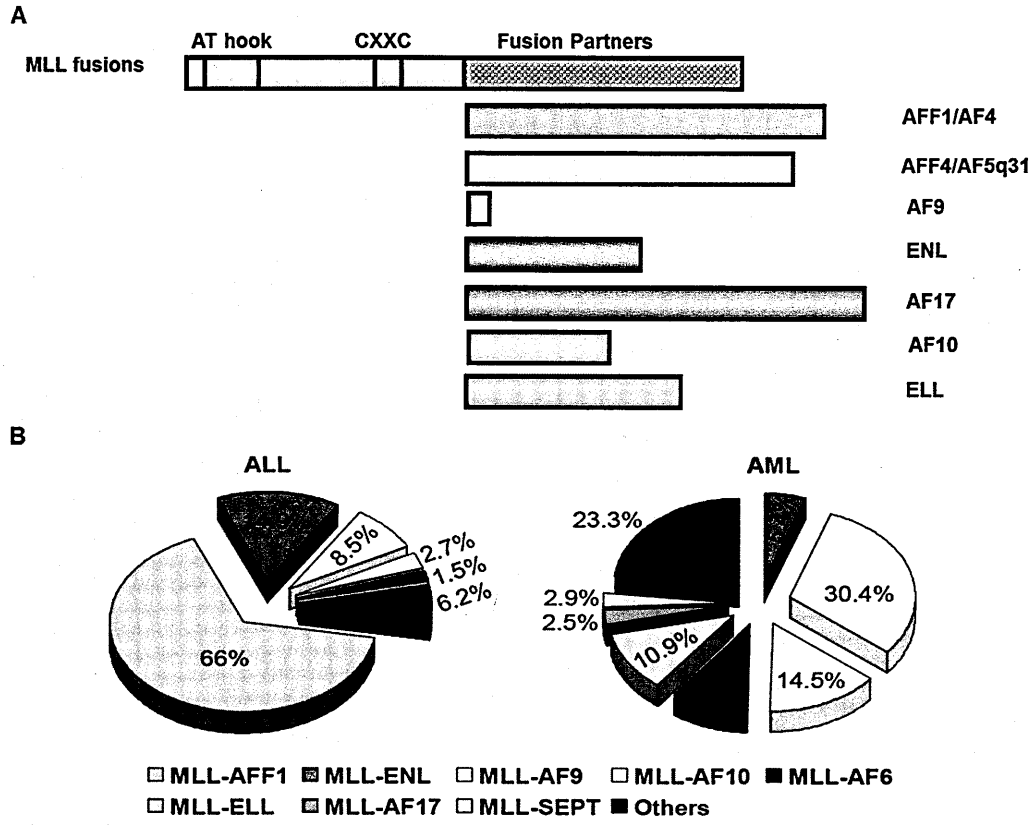


Figure 1.5. The most common MLL chimeras in acute leukemia.
(A) Several of the most frequent translocation partners are listed. (B) The distribution of MLL chimeras in Acute Lymphoblastic Leukemia (ALL) and Acute Myeloid Leukemia (AML). In MLL-rearranged ALL, *AFF1*, *AF9*, *ENL*, and *AF10* account for 90% of all the translocations, while in AML, 70% of MLL translocations involve *AF9*, *ENL*, *AF10*, *ELL*, and *AF17* (Meyer et al. 2009a). MLL-AFF1 is the most frequent MLL translocation protein.

1.2.3.1. Transcriptional Elongation Checkpoint Control (TECC) in leukemia

ELL was the first functionally characterized MLL fusion partner. The identification of ELL as a RNA Pol II elongation factor leads to the hypothesis that transcriptional elongation might play an essential role in MLL chimera-mediated leukemogenesis (Shilatifard et al., 1996) (Figure 1.6A). However, the transcriptional activation domain and the Pol II binding domain of ELL do not require the transformation activity of MLL-ELL (DiMartino et al., 2000). It is likely that the C-terminus of ELL is sufficient to recruit other transcriptional regulators to activate target genes controlled by

MLL-ELL, as demonstrated by the *Drosophila* genetic assay that the C-terminal fragment of dEll can rescue the lethality phenotype of the dEll mutant (Gerber et al., 2005b).

Besides ELL itself as an elongation factor, other nuclear fusion partners were found to associate with transcriptional elongation factors, suggesting a general role of transcriptional elongation control in MLL chimera-mediated gene activation. For example, the most common translocation partner, AF4/AFF1, and its paralog, the ALL1-fused gene from the chromosome 5q31 protein (AF5q31/AFF4), a rare fusion partner, have been shown to interact with the transcription elongation complex P-TEFb (Bitoun et al., 2007; Estable et al., 2002). AFF1 and AFF4 belong to a functionally unknown AF4/FMR2 protein family with conserved domain structures at the N- and C-terminus (Bitoun and Davies, 2005). LAF4/AFF3 (Lymphoid nuclear protein related to AF4), another member of this family of proteins, is also a MLL translocation partner. In addition, the protein-protein interaction assays identified the common translocation partners AF9/ENL as the interaction proteins of the AF4 family of proteins (Erfurth et al., 2004). Interestingly, perturbation of the interaction between MLL-AF4 and AF9 by the AF4-mimetic peptide, PFWT, inhibits the proliferation of MLL-AF4 leukemia cells, thus resulting in necrotic cell death (Palermo et al., 2008; Srinivasan et al., 2004). Therefore, it is possible that many of these nuclear fusion partners belong to a macromolecular complex containing P-TEFb. The translocation of MLL with any of the subunits might mistarget the P-TEFb complex to chromatin, resulting in the activation of its target genes.

1.2.3.2. DOT1L-mediated H3K79 methylation

DOT1L (disruptor of telomeric silencing) is the sole histone H3 lysine 79 (H3K79) methyltransferase, which plays important roles in transcriptional control(Mohan et al.,

2010b). The yeast two-hybrid experiments identified DOT1L as the interacting protein of the MLL fusion partner AF10 (Okada et al., 2005). The interaction of MLL-AF10 and DOT1L recruits DOT1L to its target genes, like *HOXA9*, resulting in the induction of H3K79 methylation (Figure 1.6B). Genome-wide studies further indicated that many of the MLL chimera target genes are marked by highly methylated H3K79 (Krivtsov et al., 2008). These preliminary data have stimulated great interest in investigating the exact role of DOT1L in MLL fusion-mediated leukemogenesis, although whether the H3K79 methylation by DOT1L is necessary and sufficient for MLL chimera-induced leukemia is still controversial.

Recent purifications of ENL-containing complexes suggested a general association of DOT1L with many of the most common fusion partners including AFF1, AF9, ENL, AF10, AF17, and AFF4 (Mueller et al., 2007). Besides DOT1L, P-TEFb was also found in this purification. However, whether the ENL-containing macromolecular complexes truly represent a single complex or whether there are actually several different ENL-containing complexes needs to be further addressed by tagging different subunits of the complexes or conventional column chromatography analysis.

1.2.3.3. Dimerization of MLL chimeras

Early studies using the MLL-GAS7 (growth arrest specific 7) and MLL-β-galactosidase (lacz) knock-in mice models demonstrated that oligomerization of MLL chimeras mediated by fusion partners is necessary and sufficient for their leukemogenic transformation (Dobson et al., 2000; So and Cleary, 2004; So et al., 2003a) (Figure 1.6C). Such a dimerization model was also observed in other MLL chimeras involving cytoplasmic fusion partners including epidermal growth factor receptor pathway substrate

15 (EPS15), gephyrin, septin 6 (SEPT6), and FK506 binding protein (FKBP) (Liedtke and Cleary, 2009). Unlike MLL-AF9 and MLL-ENL, this group of MLL chimeras induces leukemia with a longer latency (Krivtsov and Armstrong, 2007). Interestingly, like MLL-nuclear protein translocations, engineered MLL chimera carrying synthetic dimerization modules can also activate *Hoxa* gene expression (So et al., 2003a). However, the mechanism underlying how the dimerization of the MLL chimera contributes to the activation of downstream target genes remains elusive. Examination of histone modifications revealed that the activation of *Hoxa9* by the oligomerization group of MLL chimera MLL-FKBP does not induce H3K79me2, suggesting a diverse mechanism for transcriptional activation (Milne et al., 2005).

1.2.3.4. Other possible mechanisms

In addition to the potential roles of chromatin modifier DOT1L in a subset of MLL chimeras, other histone modification enzymes like E1A binding protein p300 and CREB binding protein (CBP) were directly found as MLL fusion partners (Meyer et al., 2009). Recently, it has also been reported that protein arginine methyltransferase 1 (PRMT1) is recruited to a rare MLL fusion MLL-EEN (extra eleven-nineteen leukemia fusion) targeted loci, like *Hoxa9*, mediated by SAM68 (Src-associated in mitosis of 68 kDa) (Cheung et al., 2007) (Figure 1.6D). PRMT1 directly methylates histone H4 arginine 3 (H4R3) and thus further promotes histone acetylation by p300/CBP (Cheung et al., 2007; Huang et al., 2005). Direct fusion of MLL-PRMT1 wild-type, instead of the catalytic mutant version of PRMT1, results in the transformation of primary myeloid progenitors, suggesting an essential role of protein arginine methylation in MLL fusion-induced leukemia.

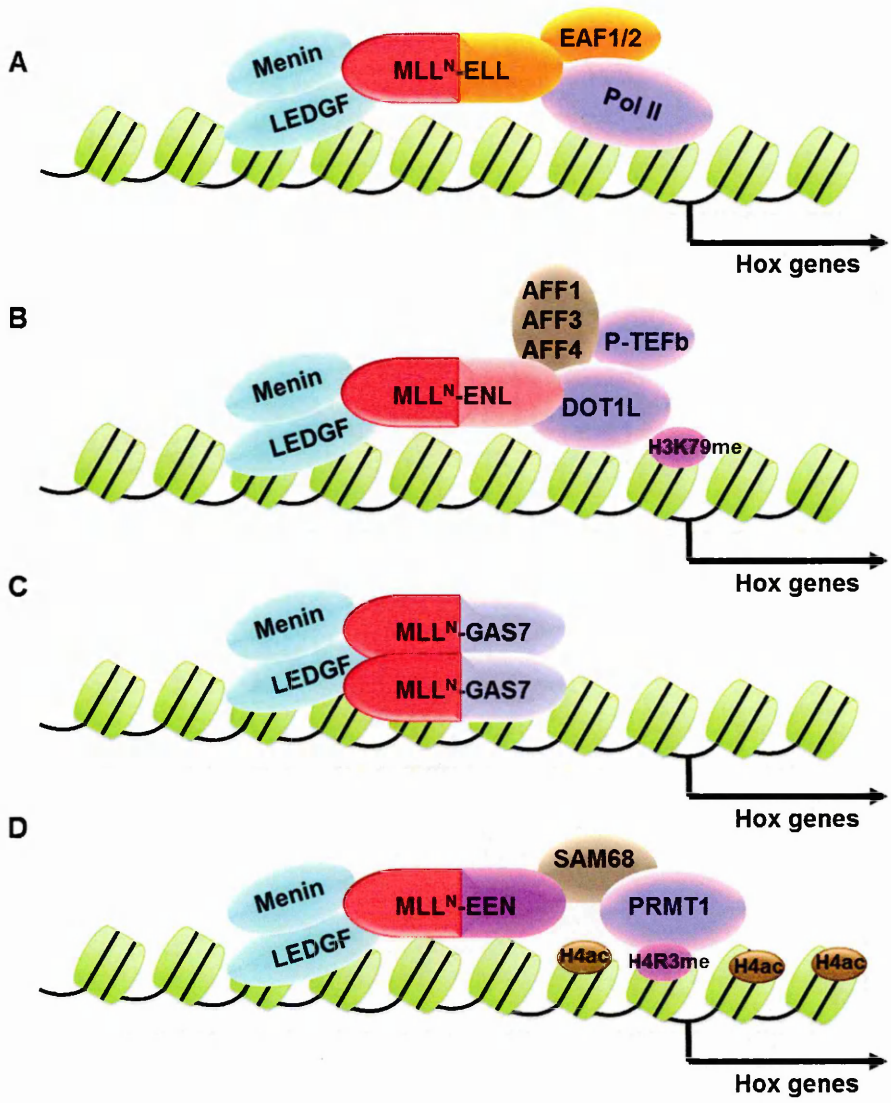


Figure 1.6. Proposed mechanisms of MLL chimera-induced leukemogenesis.
(A) Transcription elongation control. ELL was the first functionally characterized MLL fusion partner. The identification of ELL as a RNA Pol II elongation factor leads to the hypothesis that transcriptional elongation might play an essential role in MLL chimera-mediated leukemogenesis (Shilatifard et al., 1996). (B) DOT1L-mediated H3K79 methylation. The model was initially proposed based on the findings that MLL-AF10 directly interacts with DOT1L and recruits it to *HOXA9* and *HOXA10* loci, resulting in the aberrant histone H3K79 methylation. Recently, the purification of ENL-containing complexes identified many MLL fusion partners including DOT1L, leading to a plausible conclusion that these proteins exist in a single complex with DOT1L, thus putting H3K79 methylation at the center of MLL chimera induced leukemogenesis. (C) Dimerization of MLL chimeras. Several cytosol MLL fusion partners, including GAS7, contain the dimerization domain. The artificial MLL chimera protein MLL-Lacz, which dimerizes through the C-terminal Lacz domain, can also lead to leukemia. (D) Histone arginine methylation. The arginine methyltransferase 1 (PRMT1) is recruited to the MLL chimera MLL-EEN-targeted loci, like *Hoxa9*, mediated by SAM68, methylating histone H4 arginine 3. The methylated H4R3 can further stimulate the acetylation of histone H4 by p300.

1.3. Aims of my projects

Studies on MLL fusion partners have led to several hypotheses for the mechanism underlying leukemogenesis induced by MLL chimeras as mentioned above. However, it is possible that different MLL chimeras might share similar pathways for the onset of leukemia as the up-regulation of *HOXA9* and *MEIS1* are found in almost all types of leukemia with MLL translocations. The aim of this project is to explore the potential regulatory mechanism(s) responsible for the up-regulation of MLL chimera target genes. To examine if there are common factors shared by different MLL chimeras, I have directly isolated the protein complexes of several of the most frequent MLL chimeras and compared the compositions of these complexes. To further understand the roles of these factors during development and disease, I have combined ChIP and expression profile studies to analyze their downstream gene targets in mouse embryonic stem cells and leukemia cells, which will lay the foundation for the identification of key leukemic genes and serve as potential targets for leukemia treatment.

Chapter 2. Materials and Methods

Molecular biology techniques were based on the "Molecular cloning: a laboratory manual" described by Sambrook et al. (1989). Oligonucleotides were synthesized by Integrated DNA Technologies (IDT). Enzymes were supplied by New England Biolabs (NEB) and Invitrogen unless otherwise stated. Transfection reagents were obtained from Roche and Invitrogen. Standard buffer solutions and media were prepared by the Core Facility of SIMR.

2.1.Antibodies

For the homemade antibodies, antigens were expressed as Histidine-tag fusion proteins in PET-16b, purified on nickel-nitriloacetic acid (Ni-NTA)-agarose according to Qiagen's protocol and sent to Pocono Rabbit Farm and Labs for immunization into rabbits, except for dAFF4 which was injected into guinea pigs. *Drosophila* Rpb1 antibody was raised in rabbits against the synthetic peptide ERLMKKVFTDDVIKEMTDSG(C) conjugated via cysteine to *keyhole limpet* hemocyanin (KLH). The Rabbit anti-human RNA polymerase Rpb1 antibody was generated by immunization with the synthetic peptide: ERALRRTLQEDLVKDVLSNGC conjugated to KLH. All of the antibodies used in the thesis are listed in Table 2.1.

Table 2.1. List of antibodies used in this thesis.

	Homemade antibodies			**Commercial antibodies**	
Names	Rabbit No.	Antigens	Names	Cat. No.	Company
hELL1NT	4373/74	1-194	Pol II (8wg16)	MMS-126R	Covance
hELL1CT	4373/74	431-621	Pol II (N20)	sc-899	Santa Cruz
hELL2NT1	4067/68	1-297	Pol II (4H8)	ab5408	Abcam
hELL2NT2	4067/68	1-157	Pol II (H5)	MMS-129R	Covance
hELL2CT	4067/68	426-640	Pol II (H14)	MMS-134R	Covance
hELL3NT	939/940	1-248	AFF4	sc-107134	Santa Cruz
hELL3CT	941/942	255-397	CDK9	ab10874	Abcam
hAFF4NT	3592/93	76-237	Hexim1	ab25388	Abcam
hAFF4CT	3566/67	761-931	Cyclin T1	ab2098	Abcam
hAF4/AFF1NT	4063/64	73-264	pS2	ab5095	Abcam
hAF4/AFF1CT	3603/04	763-917	pS5	ab5131	Abcam
dLilliCT	2268/69	1431-1673	H3K4me2	ab7766	Abcam
dELL	308	Full length	H3K4me3	ab8580	Abcam
hCDK9CT	146/147	204-372	Histone H3	ab1791	Abcam
dRpb1		ERLMKKVFTDDVIKEMTDSG	BRD4	A301-985A	Bethyl
hRpb1		ERALRRTLQEDLVKDVLSNGC	TFIIB	SC-225	Santa Cruz
			Smc1	A300-055A	Bethyl
			Smc3	ab9263	Abcam
			MLL1-NT	A300-086A	Bethyl
			MLL1-CT	A300-374A	Bethyl

2.2. Stable cell line generation

MLL-ELL cDNA was previously described (DiMartino et al. 2000). *MLL-AF9*, *MLL-AFF1,* and *MLL-ENL* cDNAs were a gift from Dr. Jay Hess (University of Michigan). *AFF1*, *AFF2*, *AFF3*, *AFF4*, *CDK9*, and *BRD4* cDNAs were obtained from Open Biosystems. *ELL1*, *ELL2,* and *ELL3* cDNAs were previously described (Johnstone et al., 2001). Flag-tagged cDNAs were cloned into pCDNA5/FRT-TO vector (Invitrogen) modified with a N-terminal flag tag. The plasmids were then transfected into 293 Flp-in-TRex cells and selected by hygromycin. The single clones were picked and cultured up to 3 liters. The 293 Flp-in-TRex cells were grown in suspension with CD 293 medium

(Invitrogen) as described by the manufacturer or Dulbecco's Modification of Eagle's Medium (DMEM) medium with 10% fetal bovine serum (FBS).

2.3.ES Cell Culture and Differentiation

Mouse embryonic stem cells (KH2 and V6.5) were cultured on irradiated mouse embryonic fibroblast (MEF) feeder layers in 0.1% gelatin-coated tissue culture flask. Cells were grown in DMEM (D6546, Sigma) supplemented with 15% ES-certified fetal bovine serum (Hyclone), 2 mM L-glutamine, 0.1 mM nonessential amino acids (NEAA), 0.1 mM β-mercaptoethanol, and recombinant Leukemia Inhibitory Factor (LIF) (Millipore). For ChIP and RNA analysis, cells were grown for one passage of feeders on tissue culture plates for 30 minutes. Embryoid bodies (EBs) were formed by culturing 150,000/ml ES cells in ES medium without LIF (ES differentiation medium) on non-adhesive bacteriological petri dishes for indicated days (Kurosawa, 2007). Media were changed daily after two days. For neural differentiation, day-5 EBs were plated on laminin-coated 6-well tissue culture plate in ES differentiation medium with 1 mM all-trans retinoic acid (ATRA) (Sigma) for additional 14 days. Media were changed every other day. On day 14, the RA-treated EBs were immunostained by anti-βIII-Tubulin (Covance).

2.4. Lentivirus-based RNAi and Tissue Culture

Lentiviral particle preparation and infection were performed as previously described (Lin et al., 2010). Briefly, around 70% confluent 293T cells in 150 mm tissue culture plates were co-transfected with 8μg of the shRNA construct or Non-targeting control shRNA, 6 μg of PsPAX2 packaging plasmids and 2 μg of pMD2.G envelope plasmids using FuGENE 6 or X-tremeGENE 9 (Roche). The media was replaced with fresh DMEM supplemented with 10% FBS after 16 hours of transfection. The lentiviral

supernatants were collected 48 and 72 hours after the transfection, filtered through 0.45 μm filters and concentrated at 18K rpm for 2 hours.

MV4-11 cell line was a gift from Dr. Mike Thirman (U. Chicago, IL) and the Jurkat, SEM and REH cell lines were obtained from the ATCC. Kopn-8, ML-2, and EOL-1 cell lines were obtained from the DSMZ. These leukemia cell lines were grown according to the ATCC's or DSMZ's instructions. All cells were maintained at 37 °C under 5% CO_2. For lentiviral infection of MV4-11 leukemia cells, $1x10^5$ cells were seeded in RPMI 1640 media. Polybrene (Sigma) was added at a final concentration of 8μg/ml. After adding 50 μl lentiviral particles (MOI ~10), spin transduction was performed at 2000 RPM for 120 minutes at 32° C. 6 hours after infection, the media was replaced with 100 μl of RPMI 1640 media supplemented with 10% FBS and 5 ng/ml recombinant human granulocyte M-CSF (GM-CSF) (Prospect Protein Specialist). Cells were incubated at 37°C for four days before RNA extraction and RT-PCR as described for the siRNA experiments.

HCT-116 cells were grown in McCoy's 5A medium supplemented with 10% FBS. For the AFF4 and ELL2 knockdown analysis, shRNA plasmid targeting the human ELL2 mRNA (V2THS_28741 from Open Biosystem), human AFF4 mRNA (V2THS_197522 from Open Biosystem) and a non-targeting control plasmid (RHS4743 from Open Biosystem) were used. HCT-116 cells were plated in 6-well plates at 2 X 10^5 cells per well and infected with viral supernatants in the presence of 4 μg/ml of Polybrene (Sigma) for 4 hours. The infected cells were selected with 2 mM of puromycin and induced with 1 ug/ml of Doxycycline for 4 days before harvesting for Western blot analysis.

Murine V6.5 ES cells were cultured under mouse ES complete medium (Dulbecco's modified Eagle medium supplemented with 15% fetal bovine serum

(Hyclone), 1000 U/ml leukemia inhibitory factor (Millipore), nonessential amino acids, L-glutamine, Penicillin/Streptomycin and β-mercaptoethanol) on irradiated mouse embryonic fibroblasts (MEFs). For the ELL2, ELL3, Smc1a, and Smc3 knockdown analysis, shRNA plasmid targeting the mouse ELL2 mRNA (TRCN0000188411 from Open Biosystem), ELL3 mRNA (RMM4534-NM_145973), mSmc1a (RMM4534-NM_019710), and mSmc3 (RMM4534-NM_007790), and a non-targeting control plasmid (SHC002 from Sigma) were used. The lentiviral particles were resuspended in 125 ul of DMEM. For lentiviral infection of V6.5 cells, 4x105 cells were seeded in ES complete media and directly infected with 20 μl lentiviral particles in the presence of 8μg/ml of Polybrene (Sigma). 24 hours after infection, the ES cells were subjected to selection with 2 ug/ml of puromycin for an additional 48 hours.

2.5.Flag purification, MudPIT analysis, and size-exclusion chromatography

Nuclear extracts were prepared and subjected to anti-Flag agarose immunoaffinity chromatography. Trichloroacetic acid-precipitated protein mixtures from purifications were digested with endoproteinase Lys-C and trypsin (Roche) as previously described. Peptide mixtures were loaded onto triphasic 100-mm fused silica microcapillary columns as described previously. Loaded microcapillary columns were placed in-line with a Quaternary Agilent 1100 series high-pressure liquid chromatography pump and a Deca-XP ion trap mass spectrometer (Thermo Fisher) equipped with a nano-LC electrospray ionization source. Fully automated multidimensional protein identification technology (MudPIT) runs were carried out on the electrosprayed peptides. Tandem mass spectra were interpreted by using SEQUEST against a database containing *Homo sapiens* protein sequences downloaded from the National Center for Biotechnology Information. In addition to estimate false discovery rates, each sequence was randomized (keeping the

same amino acid composition and length), and the resulting "shuffled" sequences were added to the "normal" human database and searched at the same time. Peptide/spectrum matches were sorted and selected using DTASelect, keeping false discovery rates at 2% or less, and peptide hits from multiple runs were compared using CONTRAST. To estimate protein levels, spectral counts of nonredundant proteins were normalized by using the in-house-developed script NSAF7.

For the size exclusion analysis, the eluate from the Flag purification of AFF2, AFF3, and CDK9 or nuclear extracts was individually subjected to a Superose 6 HR size-exclusion chromatography column (GE Healthcare) containing size-exclusion buffer (40 mM HEPES [PH 7.5], 350 mM NaCl, 10% glycerol and 0.1% Tween-20). The fractions were analyzed by silver-staining, Western blotting, and MudPIT.

2.6. Immunoprecipitations and kinase assays

Approximately 10^7 cells for each assay were collected, washed with phosphate-buffered saline once, and lysed in high-salt lysis buffer (20 mM HEPES [pH 7.4], 10% glycerol, 0.35 M NaCl, 1 mM MgCl2, 0.5% Triton X-100, 1 mM dithiothreitol (DTT)) containing proteinase inhibitors (Sigma). After incubation at 4°C for 30 min, the lysate was cleared twice by centrifugation at 4°C. The balance buffer (20 mM HEPES [pH 7.4], 1 mM MgCl2, 10 mM KCl) was added to the resulting supernatant to make the final NaCl concentration 300 mM. The lysate was then mixed with antibodies and protein A beads or with anti-Flag agarose (Sigma). After incubation at 4°C for 4 h, the beads were spun down and washed three times with wash buffer (10 mM HEPES [pH 7.4], 1 mM MgCl2, 300 mM NaCl, 10 mM KCl, 0.2% Triton X-100) before eluting by boiling in SDS gel sample buffer. Kinase assays done as previously described (Bitouin et al. 2007).

2.7. Western Blots

Whole cell extracts from Hela, 293 T, or ES Cells were prepared by 0.5% NP-40 buffer containing proteinase inhibitors (P8340, Sigma). Proteins were resolved by SDS-PAGE gel and developed with Western Lightning Plus–ECL from Perkin. Primary antibodies used: AFF4 (1:2000), AFF1 (1:2000), ELL (1:1000), β-tubulin (Sigma-E7, 1:10000), Smc1a (A300-055A, Bethyl, 1:2000), Smc3 (ab9263, abcam, 1:2000), 8wg16 (Covance, 1:2000), H5 (Covance, 1:2000) and H14 (Covance, 1:2000). HRP-conjugated secondary antibodies from Sigma were used with a dilution of 1:5000.

2.8. Alkaline Phosphatase Characterization

Alkaline phosphatase levels were measured with the Alkaline Phosphatase substrate Kit 1 (Vector Laboratories, SK-5100) by following the manufacturer's instructions.

2.9. Chromosome Conformation Capture (3C) assay

The 3C assay was performed as previously described with minor modifications (Hagege et al., 2007). Briefly, 1 x 10^7 cells were crosslinked with 2% of paraformaldehyde at room temperature for 10 min, followed by glycine quenching and cell lysis. The nuclei were digested with HindIII overnight at 37 °C and then ligated with T4 DNA ligase at 16 °C for 4h. DNA was purified by phenol-chloroform extraction. Primer efficiencies were monitored by serial dilution. Digestion efficiencies were examined by primer pairs amplifying genomic regions spanning or devoid of HindIII sites. A bacterial artificial chromosome (BAC) containing the entire *Hoxa* locus (RP23-20F21) was digested with HindIII and religated to prepare the control template. To compare results between samples, the 3C signals were normalized to a control locus *Ercc3*.

2.10. Electron microscopy

Sperm were fixed in 4% paraformaldehyde/0.01% glutaraldehyde (PBS) and embedded in 3% gelatin. Samples were infiltrated with 2.3 M sucrose at 4°C overnight, and then mounted on aluminium stubs, frozen and sectioned. Thin sections (50 – 70 nm) were picked up in drops of 2.3M sucrose and collected on formvar and carbon coated mesh grids. After blocking in 1% BSA in PBS, the sections were incubated with primary antibodies and subsequently incubated with secondary antibody conjugated to 6 nm and/or 12 nm gold particles (Jackson ImmunoResearch Laboratories, Inc). The sections were fixed in 1% glutaraldehyde and stained with ice cold 0.4% uranyl acetate/1 % methyl cellulose (pH 4) and dried. The samples were viewed in a FEI Tecnai transmission electron microscope operated at 80kv.

2.11. Affymetrix Microarray Analysis

Affymetrix Mouse 430 v2 arrays were analyzed in R, version 2.11.1, using the packages affy (Gautier et al., 2004), version 1.26.1, and limma (Smyth et al., 2005), version 3.4.3. Normalization was done using rma. Annotation information for the probes was taken from Ensembl 62.

2.12. RNAi, RT-PCR, and Total RNA-Seq analysis

AFF1, AFF2, AFF3, and AFF4 SMARTpools from Dharmacon were used for all siRNA experiments. RNA was extracted with RNeasy from Qiagen and RNA levels were measured with QiagenSYBR green 1-step RT-PCR reagent. cDNAs were synthesized with High Capacity RNA-to-cDNA Kit from Applied Biosystems. The expression levels were measured with iQ SYBR Green Supermix from Bio-Rad on MyIQ (Bio-Rad).

Relative expression to housekeeping genes was calculated assuming 2-fold primer efficiencies.

For total RNA-seq analyses, 2.5 μg of total RNA were depleted of ribosomal RNA with the Ribo-Zero kit from Epicentre. The ribosomal RNA-depleted samples were amplified with a TruSeq RNA Sample Prep Kit (Illumina) for the further next-generation sequencing.

Reads from two biological replicates for each sample were aligned to the mouse genome UCSC mm9 and to gene annotations from Ensembl 65 using TopHat v1.4.1 (Trapnell et al., 2009). Cuffdiff v1.3.0 was used to quantify Reads Per Kilobase of transcript per Million (RPKM) values, to perform differential expression analysis at FDR < 0.05, and to assess statistically sufficient read coverage for each gene (Trapnell et al., 2010). As indicated in the figure legends, some analyses used a subset of genes, which contained the Cuffdiff status "OK" to exclude genes which were not expressed or not reliably covered to perform meaningful analysis. RNA-seq reads were not extended for track figures and are shown at single base resolution.

2.13. ChIP-qPCR and ChIP-Seq Analyses

Chromatin immunoprecipitation (ChIP) was performed according to previously described protocols (Lin et al., 2010). Briefly, cells were cross-linked with 1% paraformaldehyde and incubated with gentle rotation for 10 min at room temperature; cross-linking was quenched by the addition of glycine. Fixed cells were sonicated in lysis buffer using Bioruptor (Diagenode). Sonicated lysates equivalent to 5×10^6 cells were used for ChIP assay. ChIP products were analyzed by qPCR using SYBR green on MyIQ

thermal cycler (Bio-Rad). The comparative cycle threshold method was used to determine enrichment from replicate PCR reactions at Ell3 or Pol II-binding sites relative to the level of input. For ChIP-Seq, 5×10^7 cells were used per immunoprecipitation according to the previously described protocol (Lee et al., 2006). ChIP-Seq libraries were prepared with Illumina's ChIP-Seq sample prep kit.

Sequencing data was acquired through default Illumina pipeline using Casava v1.8. Reads were aligned to the mouse genome (UCSC mm9) using the Bowtie aligner v0.12.7 allowing uniquely mapping reads only and allowing up to two mismatches (Langmead et al., 2009). Reads were extended to 150 bases toward the interior of the sequenced fragment and normalized to total reads aligned (reads per million; RPM). The average coverage in 25 bp bins was computed across the genome and rendered in the UCSC genome browser. External sequencing data was acquired from GEO as raw reads and aligned in the same way as internally sequenced samples (uniquely mapping reads only; two mismatches; 150 bp extension length). External data tracks were also binned in 25 bp intervals for track diagrams.

Peak detection was done using MACS v1.4.1 for all samples except H3K27me3 and H3K36me3. For all samples analyzed by MACS in this study, associated control samples were used to determine statistical enrichment at a FDR < 0.05. The broad domain peak detector SICER was used to call enriched regions for H3K27me3 and H3K36me3 at the FDR < 1e-10, window size of 200, and gap size of 600.

The high-confidence enriched regions were used to depict ChIP-Seq enrichment profiles. Regions of interest are shown for each factor as a binary value of enriched/not enriched and rows were sorted by the shortest distance of an Ell3 peak to an annotated

TSS. Regions spanning 50 kb on either side of feature indicated were binned into 200 bp windows. Regions showing 5 kb on either side of the feature indicated were binned into 25 bp windows. Clustering analysis was performed using Cluster 3.0 (distance measure: Euclidean; linkage: Pairwise single-linkage; k-means:k=3) and visualized using Java TreeView. Canonical gene start sites were used for clustering analysis. Ordering for ChIP-seq enrichment profiles shown in Fig 1D was done by first annotating the nearest gene to each Ell3 peak. All peak regions were ordered by the position relative to the nearest gene and then the minimum distance of either end of an Ell3 peak to either end of the nearest gene. Ell3 regions shown are oriented 5' to 3' corresponding to the orientation of the nearest gene. The 'S'-curve profiles shown are ordered by the position of an Ell3 peak region relative to its nearest gene (downstream, TSS/inside, upstream), and then sorted by the distance to the gene. Fig 2A was ordered based on the position and minimum distance of each Ell3 nearest gene to the nearest Ell3 peak. The enriched ChIP-Seq signals for Ell3, Pol II and the histone modifications are shown within 50 kb around the TSS of these genes. Each line represents a gene, and color indicates enrichment. Each cluster was individually sorted based on the position and minimum distance of Ell3 to the nearest TSS. Each cluster was ordered independently. Gene regions are shown 5' to 3'.

Gene annotations and transcript start site information were from Ensembl 65. Bound genes were called if an enriched region of the factor occurred within 1kb of the start site of any isoform of the gene. GO analysis was performed using the regions indicated in GREAT v2.0.1 or with the gene list indicated at DAVID (accessed April 4 2012).

Data generated for this study (GEO accession number GSE38148) includes ChIP-seq data for Ell3, Pol II_NonT_shRNA and Pol II_Ell3_shRNA (the two Pol II samples were used for Pol II gain/loss analysis). GSE38148 also includes RNA-seq data generated

for this study: ES_NonT_shRNA, ES_Ell3_shRNA, and EB5_NonT_shRNA. Other data sets come from previously published studies. p300, H3K4me1 and H3K27ac ChIP-seq data are from GEO accession number GSE24164 (Creyghton et al., 2010). Nipbl, Smc1, Smc3, Med1, Med12 and Ctcf ChIP-seq data are from GEO accession number GSE22557 (Kagey et al., 2010). ChIP-seq data for Ell2, Aff4, as well as the Pol II that was used in track diagrams and for clustering, comes from GEO accession number GSE30267 (Lin et al., 2011). Oct4 and H3K36me3 ChIP-seq data comes from GEO accession number GSE11724 (Marson et al., 2008). H3K4me3 and H3K27me3 ChIP-seq data are from GEO accession number GSE12241 (Mikkelsen et al., 2007). Ell ChIP-seq data are from GEO accession number GSE32120 (Smith et al., 2011b).

2.14. Track Figures

Read coverage information in the track figures was created using R by extending the reads 150 bases toward the interior of the sequenced fragment and then by computing the number of extended reads in 25 bp windows as the count of extended reads per million reads sequenced (RPM; counts/million). The resulting coverage object was exported and visualized using the UCSC genome browser (Kent et al., 2002).

2.15. Histogram and Heatmap Figures

Histogram representations of ChIP-seq binding for Pol II and SEC-members were analyzed using R. First, all gene annotations and enriched peak regions were loaded. For each gene region, +/- 5 kb surrounding the transcription start site was calculated. Using 50bp windows tiling the 10 kb regions, enriched peak regions were used to label a tile either enriched or not enriched. The resulting data structure contained 200 columns, the number of rows equals the number of annotated genes in the genome, and a one or zero in each position of the matrix indicating enrichment.

The heatmap representation of the microarray expression values was also done in R using all probes that had at least a two-fold change in expression, up or down, at six hours of induction versus no induction. For each time-point and replicate depicted (2, 4, and 6 hour), expression values were converted to fold-changes relative to the 0 hour (wild-type) time-point. Log2 fold-changes were then binned into nine equally spaced groups from greater than 2 to less than -2, in 0.5 value increments. The three replicates for all of the three time-points were combined into a matrix, and then sorted based on the total sum of bin magnitudes.

2.16. Pol II occupancy Analysis

The Pol II ChIP-seq analysis comparing non-targeting shRNA and Ell3 shRNA treated samples were rank normalized as described (Rahl et al., 2010). Using 25bp bins tiling the mouse genome (UCSC mm9), we computed the total extended read (fragment) coverage for each bin, sorted the bins by greatest to least fragment coverage, and replaced the coverage value for both sample bins at the same rank as the mean value of coverage between the samples (NonT shRNA and Ell3 shRNA). Thus, each bin in the genome was given a value of rank normalized counts. For all Pol II bound genes, the isoform's start site with the maximum Pol II occupancy was selected as the TSS of that gene. Metagene plots in Fig. 3D show 5kb on either side of the TSS as the average rank normalized counts for the Pol II-bound genes within gene subset indicated. All transcription start sites for each nearest gene were computed.

Chapter 3. The Identification of the Super Elongation Complex (SEC) and its role in leukemogenesis

3.1. Introduction

The MLL translocation-based leukemia involves a large number of fusion partners, many of which share little sequence or known functional similarity. Recent studies showed that the nuclear subgroup proteins, AF4, AF9, AF10, and ENL may exist in the same complex and interact directly or indirectly with H3K79 methyltransferase, DOT1, leading to the suggestion that Dot1-mediated methylation of H3K79 was central to leukemogenesis in patients with MLL translocations (Bitoun et al., 2007; Krivtsov et al., 2008; Mueller et al., 2007; Mueller et al., 2009; Okada et al., 2005; Zhang et al., 2006). However, at this time, there is little evidence, and no mechanistic understanding, for how H3K79 methylation by Dot1 could lead to gene activation.

The report of ELL1 being a Pol II elongation factor was the first biochemical and molecular characterization of any of the MLL partners in leukemia, and to date, ELL1 is the best functionally characterized of the MLL partners(Shilatifard et al., 1996). It has been postulated that perhaps other MLL partners may also function in the regulation of transcription elongation as well (Shilatifard et al., 2003). Furthermore, ELL1, one notable translocation partner of MLL, which is a focus of my studies and has a demonstrated role in transcription elongation, was not reported to be a part of the Dot11 complexes (Mueller et al., 2007; Mueller et al., 2009). In this chapter, I will show the biochemical search for the identification of commonalities in the disparate MLL-fusions by epitope tagging some of the most common MLL fusion partners.

3.2.Results

3.2.1. AFF4 is a shared subunit of several common MLL fusion protein complexes.

In order to begin to understand how the misregulation of gene expression is caused by MLL-fusion proteins, I expressed some of the most common MLL fusion proteins, MLL-ELL1, MLL-ENL, MLL-AFF1, and MLL-AF9, in 293 cells with a Flag epitope tag under an inducible promoter, each integrated at the same site within the genome. Expression and purification of the MLL-N-terminal region most frequently found in MLL-fusion proteins resulted in the isolation of Menin, which is known to associate with the N-terminus of MLL and LEDGF, an interactor of Menin (Table 2)(Yokoyama and Cleary, 2008). Following the biochemical isolation of MLL-ELL1, MLL-AFF1, MLL-AF9, and MLL-ENL (Figure 3-1A), these purified complexes were subjected to Multidimensional Protein Identification Technology (MudPIT) to carry out proteomic analyses for each complex. While the MLL-AF9 and MLL-ENL identified a few of the proteins previously described as ENL-associated proteins, including AFF1, AFF4, and Dot1 (Mueller et al., 2007; Mueller et al., 2009), the MLL-ELL1 and MLL-AFF1 chimera complexes included AFF4, but notably, not Dot1 (Table 2). In fact, AFF4, which itself is a rare translocation partner of MLL, is a shared

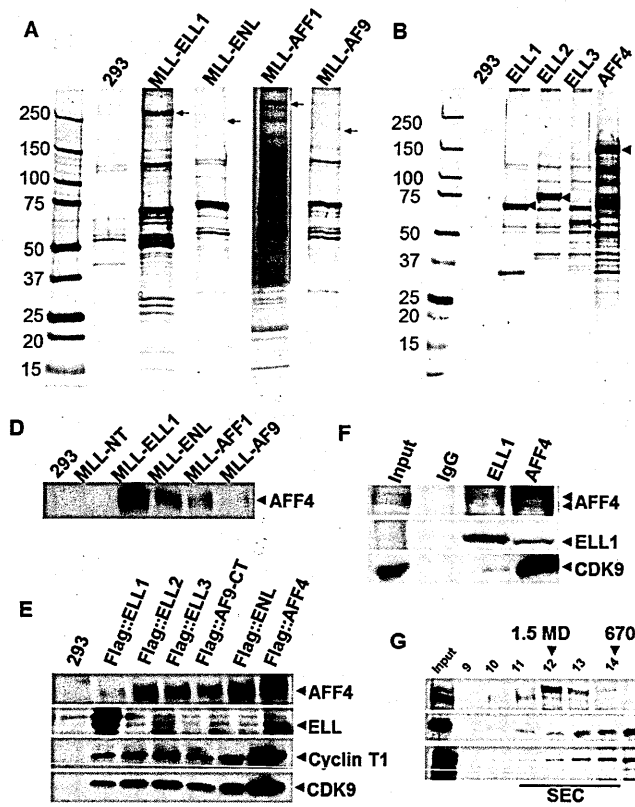
Table 3-1. MudPIT analysis of MLL chimera purifications.

Protein	MLL-NT dNSAF AVG	MLL-ELL1 dNSAF AVG	MLL-AFF1 dNSAF AVG	MLL-AF9 dNSAF AVG	MLL-ENL dNSAF AVG
MLL-NT	0.050949	0	0	0	0
MLL-ELL1	0	0.084755	0	0	0
MLL-AFF1	0	0	0.012444	0	0
MLL-AF9	0	0	0	0.051766	0
MLL-ENL	0	0	0	0	0.06576
Menin	0.21546	0.088819	0.04613	0.053286	0.151354
LEDGF	0.000885	0.000533	0.000192	0.001115	0.001327
AFF4	0	0.001971	0.000831	0.000593	0.000127
Dot1L	0	0	0	0.001484	0.000048

subunit for all of the purified MLL chimeras (Table 2).

3.2.2. AFF4 forms the Super Elongation Complex (SEC) with ELLs and P-TEFb.

Given the unexpected observation of finding this largely uncharacterized protein associating with the MLL chimeras, I, therefore, generated a cell line expressing Flag epitope-tagged AFF4 and identified associated proteins (Figure 3.1B-C). Surprisingly, all three ELL proteins were found in the isolated complexes. In turn, expressing and isolating Flag-tagged ELL1, ELL2, and ELL3 revealed AFF4 associating with each ELL (Figure 3.1B-C). Furthermore, the ELL and AFF4-containing complexes also consist of additional MLL partners, AFF1, ENL, and AF9 (Figure 3.1C). Another subunit of the AFF4 – ELL1 complex is the component of the Pol II C-terminal domain (CTD) kinase, P-TEFb, consisting of Cdk9 with cyclin T1, T2a, or T2b (Figure 3-1C). I also detected the previously identified ELL-associated factors, EAFs (Simone et al., 2003; Simone et al., 2001), in the AFF4, ELL2, and ELL3 complexes (Figure 3.1C). Since the purification of the ELL1 complexes were performed with ELL1 lacking the first 50 amino acids (missing in the MLL-ELL1 chimera and required for interactions with EAFs), I did not detect any of the EAFs in the ELL1 purifications. Given the fact that EAFs enhance the *in vitro* transcription elongation properties of ELLs (Kong et al., 2005), it is interesting that I also observe these factors with the AFF4-containing complexes.



C

AFF4	ELL1	ELL2	ELL3
ELL1	ELL1	ELL1	ELL1
ELL2	ELL2	ELL2	
ELL3			ELL3
AFF1	AFF1	AFF1	AFF1
AFF4	AFF4	AFF4	AFF4
AF9	AF9	AF9	AF9
ENL	ENL	ENL	ENL
CDK9	CDK9	CDK9	CDK9
Cyclin T1	Cyclin T1	Cyclin T1	Cyclin T1
Cyclin T2a	Cyclin T2a	Cyclin T2a	Cyclin T2a
Cyclin T2b	Cyclin T2b	Cyclin T2b	Cyclin T2b
EAF1		EAF1	EAF1
		EAF2	EAF2

Figure 3.1. AFF4 is a shared subunit of several of the MLL chimeras and associates with known RNA polymerase II elongation factors.

(A) Flag-tagged MLL-ELL1, MLL-ENL, MLL-AFF1, and MLL-AF9 were purified using the FLAG-affinity purification method and analyzed by silver staining and mass spectrometry. Arrows indicate the position of the Flag-tagged proteins. (B-C) Purification of ELL1, ELL2, ELL3, and AFF4 complexes. Arrow heads in (B) indicate the position of the Flag-tagged subunit. ELL1 and its paralogs, ELL2 and ELL3, were separately purified and demonstrated a similar set of associated proteins as found in the AFF4 purification. (C) AFF4 (dark red) was found in all of the Flag-ELL purifications indicating that it is a component of a novel RNA polymerase II elongation complex. (D-F) Confirmation of an interaction of AFF4 with the MLL chimeras and components of the P-TEFb elongation factor by Flag and/or endogenous immunoprecipitations. Arrowheads show the position of the protein probed by Western analysis. (D) Flag immunoprecipitations of MLL chimeras demonstrate an association of AFF4 with all chimeras, but not with a Flag-tagged MLL-N-terminal domain common to all chimeras. (E) Western blot analyses of ELL1, ELL2, ELL3, AF9, ENL, and AFF4 immunoprecipitations confirm the observed interactions of Cyclin T1 and CDK9 with these factors. (F) The endogenous association of P-TEFb with AFF4 and ELL1. (G) Size exclusion chromatography of HeLa nuclear extracts. Fractions corresponding to 1.5 MDa and 670 kDa are indicated with arrowheads and referred to as the Super Elongation Complex (SEC). The MudPIT analysis was done by the collaboration with Michael, Laurence, and Selene from the Proteomics Center.

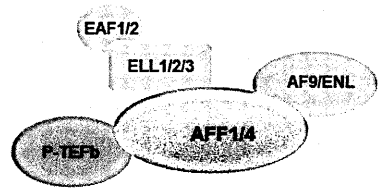
The observed interactions between ELL1-3, AFF4, and the components of P-TEFb were also confirmed by Flag and endogenous co-immunoprecipitations (Figure 3.1 D-F). In different preparations and with different tagged subunits, the relative amounts of some subunits in the isolated complexes can vary (e.g. see ELL1 and AFF4 levels in Figure 3.1 D-E). Therefore, it is important to tag and purify multiple subunits to get a clear picture of the complexes *in vivo*. Indeed, previous interpretations that Dot1, AFF1, and AFF4 exist in a single complex were primarily based on these proteins co-purifying with a single subunit, ENL (Mueller et al., 2007; Mueller et al., 2009). I also find that ENL associates with AFF1, AFF4, and Dot1, but importantly, I find that Dot1 is not associated with AFF1, AFF4, or the ELL complexes indicating that ENL is part of at least two distinct complexes (Table 3-2).

Table 3-2. MudPIT analysis of ELL1-3 and AFF4 purifications.

Protein	ELL1 dNSAF AVG	ELL2 dNSAF AVG	ELL3 dNSAF AVG	AFF4 dNSAF AVG
ELL1	0.185148	0.000385	0.003368	0.006379
ELL2	0.000542	0.054255	0	0.004369
ELL3	0	0	0.111228	0.006603
AFF4	0.004663	0.006674	0.015106	0.03917
AFF1	0.001422	0.001296	0.003659	0.00077
AF9	0.001334	0.00052	0.002346	0.005846
ENL	0.001637	0.000988	0.003198	0.011151
CDK9	0.005979	0.002635	0.007835	0.029597
CCNT1	0.00124	0.000513	0.001316	0.021903
Cyclin T 2a	0.000261	0.000071	0.001236	0.003648
Cyclin T 2b	0.000237	0.000065	0.001122	0.003313
EAF1	0	0.072906	0.02451	0.000652
EAF2	0	0.002226	0.002792	0

To further characterize the AFF4-containing complexes, I analyzed nuclear extracts by their application to size exclusion chromatography, followed by SDS-PAGE and Western analysis of AFF4 and the components of the P-TEFb elongation complex (Figure

3.1G). These studies clearly indicate that a small portion of P-TEFb co-purifies with the AFF4 large complex at about 1.5 MDa (Figure 3.1G, fraction 11-13), which I call the Super Elongation Complex (SEC) due to the presence of multiple Pol II elongation factors (Figure 3.2). Overall, these studies reveal that many of the MLL partners found in leukemia, which have very little sequence or seemingly functional similarities, are found in large macromolecular complexes associated with the Pol II elongation factors ELL and P-TEFb.



Super Elongation Complex
(SEC)

Figure 3.2. Schematic representation of the Super Elongation Complex (SEC).
The SEC is a P-TEFb-containing complex that contains various combinations of four types of proteins: the ELL1-3; EAF1-2; AFF1 and AFF4; and AF9 and ENL. P-TEFb itself consists of Cdk9 and CycT1/2 and is best characterized as a RNA Pol II C-terminal domain (CTD) kinase.

3.2.3. SEC is one of the most active P-TEFb-containing complexes.

P-TEFb is a CTD kinase involved in the regulation of transcription elongation by Pol II and can exist in both active and inactive forms (Peterlin and Price, 2006). To determine whether the purified ELL and AFF4-containing complexes contain active P-TEFb, I tested the kinase activity of these purified complexes towards the GST-Pol II C-terminal domain fusion protein (GST-CTD). The ELL1, ELL2, ELL3, and AFF4 complexes were assayed in the presence and/or absence of ATP and the GST-CTD (Figure 3.3A). The resulting products were subjected to SDS-PAGE followed by Western analysis with antibodies specific to Pol II CTD either phosphorylated on serine 2 (pS2) or serine 5 (pS5) (Figure 3.3A). I also tested the autophosphorylation of CDK9 and the possible

phosphorylation of AFF4 by P-TEFb (Figure 3.3A). From these studies, it appears that the purified ELL and AFF4-containing complexes are active as a Pol II CTD kinase. These results also suggest that AFF4 is phosphorylated by P-TEFb (Figure 3.3A), a phenomenon observed previously for AFF1 (Bitoun et al., 2007). Sequence alignment further demonstrates similarly repeated Serine-Proline (SP) motifs among AF4 family proteins (Figure 3.3B). Similar CTD kinase activities are found in ELL1 and MLL-fusion protein complexes (Figure 3.3C).

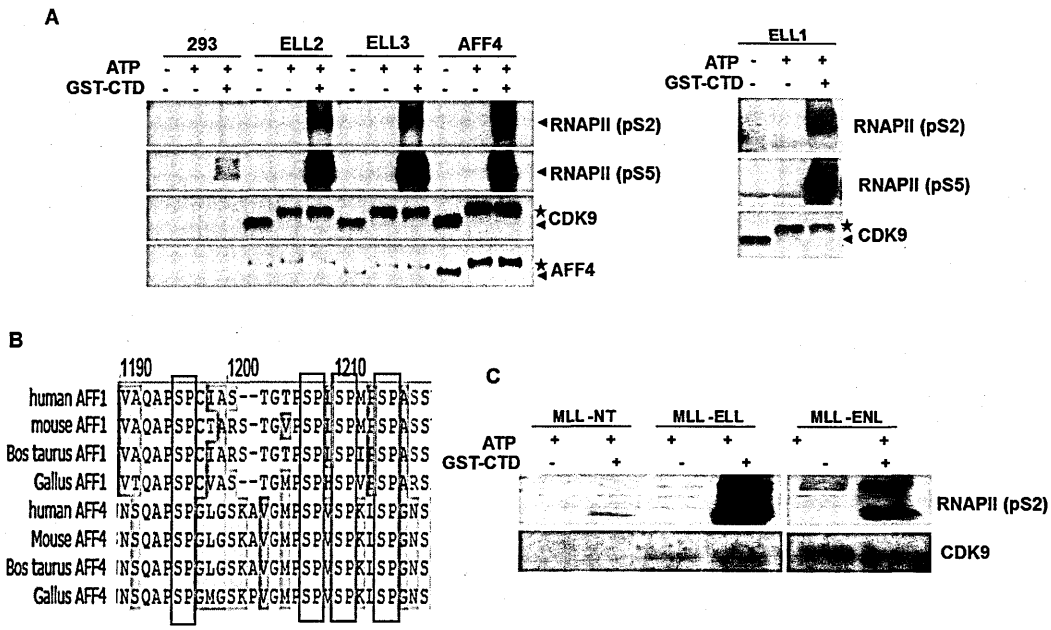


Figure 3.3. SEC can phosphorylate Pol II CTD *in vitro*.

(A) Pol II C-terminal domain (CTD) Kinase assays with the ELL1, ELL2, ELL3, and AFF4-containing complexes were performed with GST-Pol II C-terminal domain fusion protein (GST-CTD). ELL1, ELL2, ELL3, or AFF4 complexes were assayed in the presence of ATP and/or the GST-CTD and subjected to Western blot analyses with antibodies specific to Pol II CTD phosphoserine 2 (pS2), phosphoserine 5 (pS5), CDK9, and AFF4. Consistent with previous observations, serine 2 and serine 5 of Pol II CTD are good substrates for the P-TEFb complexes *in vitro*. CDK9, itself, is also known to be auto-phosphorylated, resulting in a shift in gel migration in SDS- PAGE (indicated by star, while an arrow indicates the faster migrating unphosphorylated form). AFF4 shows a similar gel mobility shift as CDK9, also indicated by an asterisk, suggesting that it is a substrate for P-TEFb as well. See Figure S1 for additional kinase assays. (B) Sequence alignment of a potential site for multiple phosphorylation of AFF4-related proteins bearing SP motifs favored by P-TEFb. (C) MLL-NT, lacking a fusion partner, MLL-ELL1 and MLL-ENL were incubated with ATP and in the presence or absence of GST-CTD and probed for CTD phosphorylation on Serine 2.

Multiple P-TEFb-containing complexes exist in mammals, but they function differently according to their regulatory subunits, such as HEXIM1/7SK, BRD4, and AFF4 (Luo et al., 2012b). In order to compare the kinase activities of these different P-TEFb-containing complexes, we fractionated the affinity-purified CDK9 complexes by size-exclusion chromatography (Figure 3.4A) and analyzed the fractions by western blot and MudPIT. Each fraction was first titrated by Western blotting to determine the level of the CDK9 protein to ensure that similar amounts of CDK9 were used to perform Pol II CTD kinase assays (Figure 3.4A). The reaction mixtures were then subjected to SDS-PAGE followed by autoradiography. Fractions 11-13, which contain SEC, (Figure 3.4A), show the strongest kinase activities *in vitro* towards Pol II CTD (Figure 3.4B, please compare activities in fractions 11-13 to fractions 16-18). Overall, although a small amount of the P-TEFb is found in these large SEC-containing complexes, the majority of the CTD kinase activity of P-TEFb is associated in these fractions suggesting that the most active forms of P-TEFb are found within the SEC family.

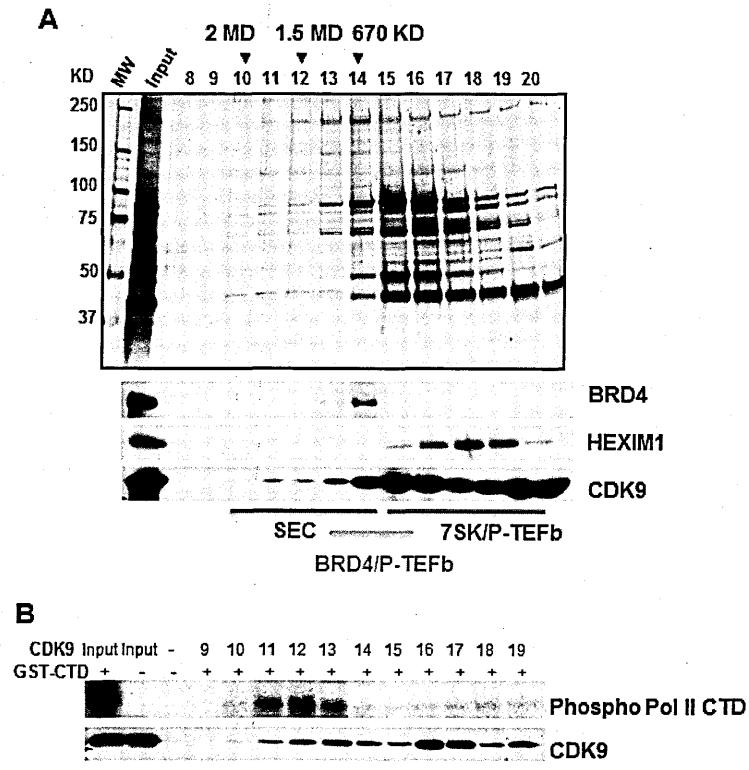


Figure 3.4. The kinase activities of different P-TEFb-containing complexes.
(A) Purification and size exclusion chromatography of P-TEFb complexes. P-TEFb complexes were isolated from Flag-CDK9 expressing HEK293T cells by Flag purification. Size exclusion chromatography was used to separate different P-TEFb complexes including SEC (peak from Fraction 10 to 14); BRD4/P-TEFb complex (peak in Fraction 14 and 15); and the HEXIM1/7SK/P-TEFb complex (peak from Fraction 15 to 19). The fractions were analyzed by silver staining and western blot. (B) Pol II CTD kinase activity analyses of the fractionated P-TEFb complexes. The amount of each fraction used was adjusted to ensure that similar amounts of CDK9 were present in each assay with γ-32P-ATP and recombinant Pol II CTD. The reaction was then subjected to SDS-PAGE and autoradiography to assess the phosphorylated Pol II levels in each reaction. This figure was done by the collaboration with Dr. Zhuojuan Luo and Nima.

3.2.4. AFF4 is a central component of SEC.

To determine which of the components of the AFF4 complex are required for complex stability and association with the P-TEFb kinase, I reduced the levels of several components of the complex using RNAi (Figure 3.5A-B). I observed that the reduction of the AFF4 homologue AFF1 does not alter ELL1 and P-TEFb stability in these cells (Figure 3.5B). However, the loss of AFF4 results in the instability of ELL1 with no significant effect on the stability of the P-TEFb components (Figure 3.5B). My studies so far indicate

that the AFF4-containing complex associates with a small portion of the P-TEFb components either when the AFF4 complex is purified (Figure 3.1B-F), or when nuclear extracts were analyzed by size exclusion chromatography (Figure 3.1G). I, therefore, tested for the association of P-TEFb with the AFF4/ELL complex in the absence of AFF4 (Figure 3.5C). Nuclear extracts from cells treated with AFF4 RNAi were subjected to size exclusion chromatography and the fractions were analyzed by SDS-PAGE, followed by Western analyses using antibodies specific to Pol II, CDK9, Cyclin T1, and Hexim 1 (Figure 3.5C). This biochemical analysis demonstrated that reduction in AFF4 levels results in the loss of association of CDK9 and Cyclin T1 with the large AFF4-containing complex (Figure 3.5C fractions 11-13), but not the Hexim1-containing P-TEFb complexes (Figure 3.5C) (Peterlin and Price, 2006). Together, these results demonstrate that AFF4 is a central component of the P-TEFb/ELL complexes.

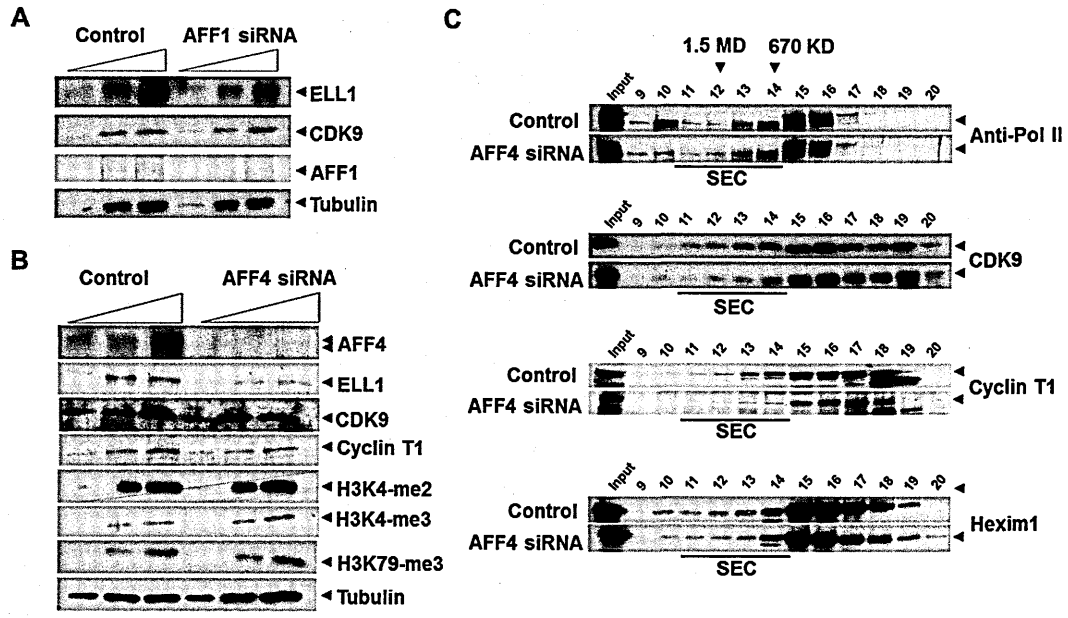


Figure 3.5. AFF4 is required for the assembly of SEC-containing ELLs, P-TEFb, and MLL partners. (A-B) AFF4, and not AFF1, is required for stability of the large elongation complex containing ELL1, P-TEFb, and MLL-partners in HeLa cells. Western blot analysis of ELL1, CDK9, and Cyclin T1 was performed in the presence and absence of AFF1 or AFF4. Nuclear extracts from the siRNA-mediated for AFF4 or AFF1 were analyzed by SDS-PAGE and Western blot analysis. Arrows indicate increasing protein loads. Bulk protein levels of ELL1 are reduced in AFF4, but not AFF1 knockdown in these cells. Bulk protein levels of P-TEFb were not affected by AFF1 or AFF4 RNAi. Global H3K4 and H3K79 methylation levels were not affected by AFF4 knockdown. Tubulin serves as a loading control. (C) Gel filtration analyses of nuclear extracts from control and AFF4-directed siRNA treated cells. Larger P-TEFb-containing complexes (fractions 10-14 also seen in Figure 1G and indicated by underlining in red) are reduced in AFF4 knockdown cells, indicating that the presence of AFF4 is required for the assembly of this complex.

3.2.5. SEC functions as a transcriptional elongation complex.

The first *in vivo* characterization of ELL as a transcription elongation factor was in *Drosophila*, where there is only one ELL-like protein, dELL (Eissenberg et al., 2002; Gerber et al., 2001). Indeed, the first hint of a connection between P-TEFb and ELL was the RNAi-mediated knockdown of Cdk9 and the loss of dELL from chromatin (Eissenberg et al., 2007). Additionally, although it took years to identify the affected genes, dELL and the sole *Drosophila* homologue of AFF4 (Figure 3.6) were part of a small set of genes isolated in a screen for Ras signaling components (Eissenberg et al., 2002; Neufeld et al.,

1998; Su et al., 2001; Tang et al., 2001; Wittwer et al., 2001). Based on our new findings, I was interested to extend these intriguing links between ELL, P-TEFb, and AFF4 in *Drosophila*.

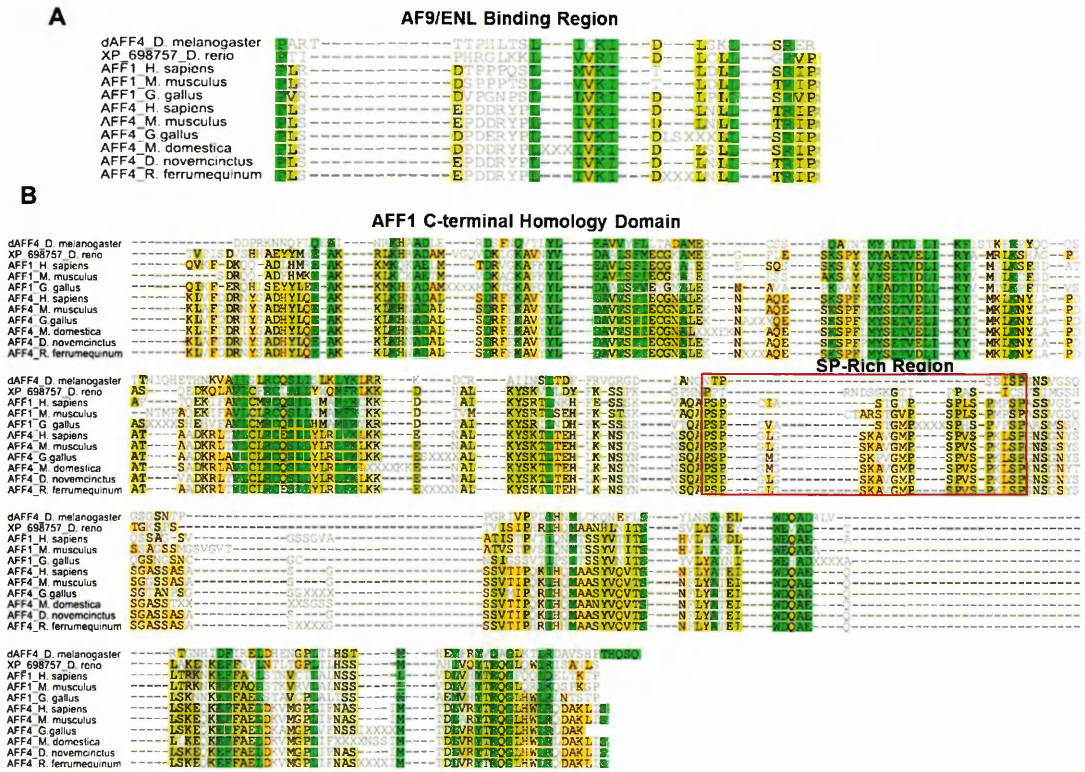


Figure 3.6. Alignment of the AFF1 and AFF4 proteins with *Drosophila* AFF4 (dAFF4).
(A) AF9 and ENL binding region of AFF1/AFF4 (Erfurth et al., 2004). (B) AFF1 C-terminal homology domain alignment of AFF1 and AFF4 from several vertebrates with the sole *Drosophila* member of this family, dAFF4, encoded by the *lilli* gene (Su et al., 2001; Tang et al., 2001; Wittwer et al., 2001). The blue boxed region shows the SP rich region from Fig. 2B that contains potential phosphorylation sites for P-TEFb.

Since many *Drosophila* Pol II elongation factors have been shown to associate with elongating Pol II on chromatin and relocalize to heat shock loci upon stress (Ardehali et al., 2009; Eissenberg et al., 2002; Gerber et al., 2001; Gerber et al., 2005a; Gerber et al., 2005b; Smith et al., 2008; Tenney et al., 2006), we generated polyclonal sera to dAFF4 and performed colocalization studies of dAFF4 with dELL and the elongating form of Pol

II. dELL and the elongating form of Pol II colocalize extensively with dAFF4 (Figure 3.6A-B), not seen with preimmune sera (data not shown). The *Hsp70* loci in *Drosophila* have been used as a model system for studying transcription elongation. The *Hsp70* loci contain poised polymerase, which upon heat shock is phosphorylated at Serine 2 in the CTD repeats by P-TEFb, allowing productive transcription (Boehm et al., 2003). We assayed the presence of dAFF4 at *Hsp70* after heat shock and observe that indeed dAFF4 colocalizes with dELL and the elongating form of Pol II on polytene chromosomes at major heat shock loci, including the *Hsp70* genes at 87A and 87C (Figure 3.7C-E). Chromatin immunoprecipitation of dAFF4 shows that it becomes associated with *Hsp70* upon heat shock, and is present throughout the transcribed unit (Figure 3.7F), similar to what was previously observed for P-TEFb (Boehm et al., 2003).

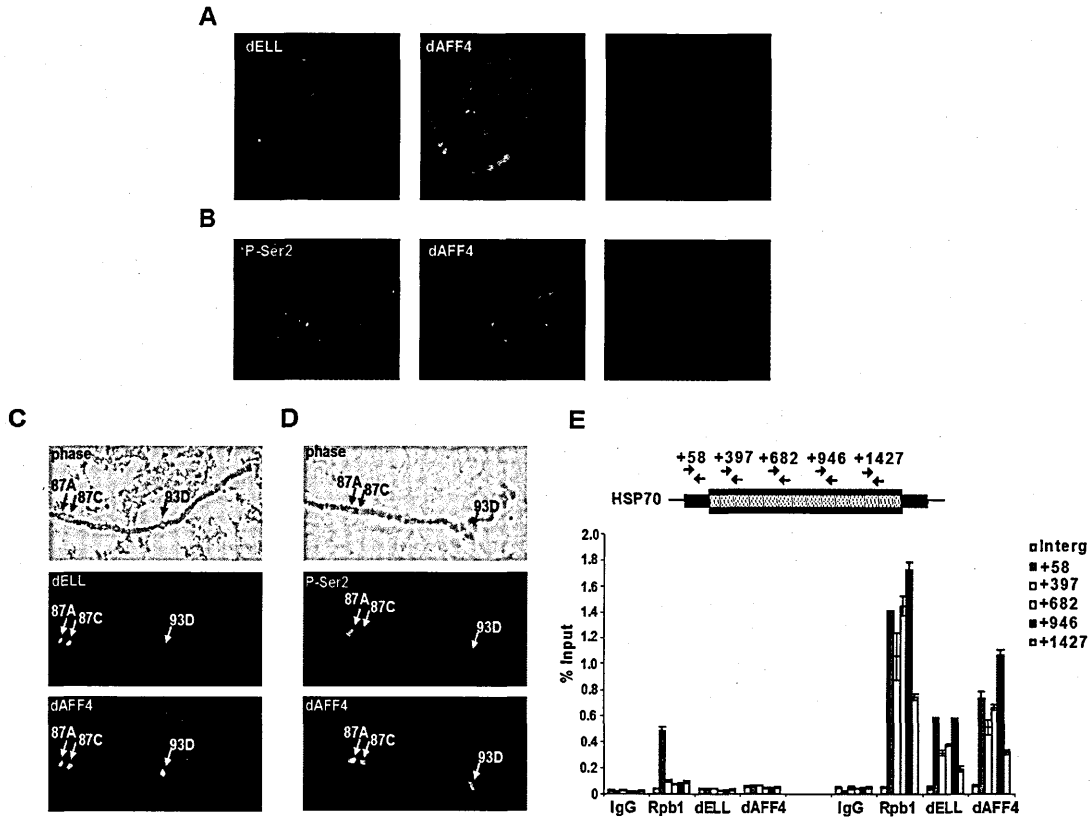


Figure 3.7. The *Drosophila* ortholog of AFF4 colocalizes with ELL and the elongating form of Pol II on *Drosophila* polytene chromosomes.

(A and B) Polytene chromosome preparations from 3rd instar larval salivary glands were probed with antibodies to dELL (A, red) or the H5 monoclonal antibody recognizing the Ser-2 phosphorylated (P-Ser2), elongating form of Pol II (B, red). Both antibodies show substantial colocalization with dAFF4 (A and B, green). Chromosomes were counterstained with DAPI (A and B, blue). (C and D) Polytene chromosomes were prepared from heat shocked 3rd instar larvae and stained as in (A and B). Phase contrast images show positions of the 87A, 87C, and 93D major heat shock loci. dAFF4 is recruited along with dELL at these loci after heat shock, associated with the P-Ser2 form of RNA Pol II. See Figure S4 for additional images. (E) Chromatin immunoprecipitation of dAFF4 and RNA Pol II large subunit (Rpb1) at *Hsp70* before and after 10 minutes of heat shock at 37° C. While Rpb1 is present at *Hsp70* prior to heat shock, dAFF4 can only be detected at *Hsp70* after heat shock, where it is found throughout the transcription unit along with Pol II. Hsp70 primers have been previously described (Boehm et al., 2003). Error bars represent standard deviations. This figure was performed by Dr. Smith Edwin.

Using chromatin immunoprecipitation, human AFF4 levels were also measured across the *HSP70* gene before and after heat shock in HeLa cells (Figure 3.8A and 3.5B). Upon heat shock, AFF4 is found at the *HSP70* promoter and throughout the transcribed

region along with RNA Pol II (Figure 3.8B-D). Interestingly, ELL2 is also recruited to the 5' end of *HSP70*, but are not significantly enriched as far into the body of the gene as AFF4 (Figure 3.8E). This could reflect different sensitivities of the antibodies employed or conceivably to differential usage of the elongation factors in this complex at distinct steps of transcription elongation.

The effect of AFF4 recruitment to *HSP70* was assessed by siRNA-mediated knockdown of AFF4. Knockdown of AFF4 leads to a defective heat shock response, showing reduced induction of *HSP70* compared to control siRNA-treated cells (Figure 3.8F). While *HSP70* is used as a model gene for studying transcription elongation, I recognize that other factors, not known to directly stimulate transcription elongation, also travel with the polymerase, such as components of the exosome (Andrulis et al., 2002). However, based upon the proven *in vitro* stimulation of transcription elongation by ELL1, the requirement for AFF4 in the stability of the P-TEFb-AFF4-ELL complex, the association of AFF4 with *HSP70* upon heat shock, and its requirement for the full expression of *HSP70*, I propose that AFF4 is a central component of the SEC complex.

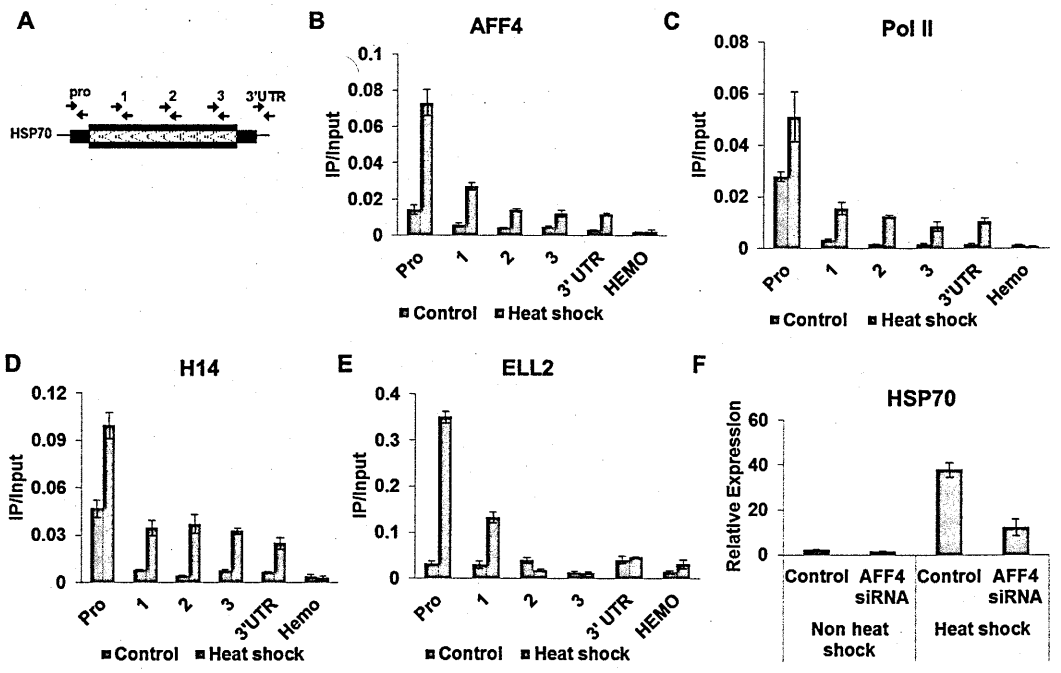


Figure 3.8. AFF4 is required for proper *HSP70* induction.

HeLa cells were heat shocked by incubation at 42° C for 2 hours. Non-heat shocked and heat-shocked cells were used in chromatin immunoprecipitation-quantitative PCR (ChIP-qPCR) assays with AFF4, ELL2, general Pol II, and the H14 monoclonal antibody recognizing the Serine 5 phosphorylated form of Pol II (B-F). (A) Position or primer pairs used for QPCR along the *HSP70* gene are indicated. (B-E) AFF4 is recruited to the *HSP70* gene after heat shock along with ELL2 and RNA polymerase II. (F) Knockdown of AFF4 in HeLa cells by RNAi inhibits *HSP70* induction. Control and *AFF4* siRNA-treated cells were heat shocked as in (A) and *HSP70* mRNA levels were assessed by quantitative RT-PCR and normalized to *GAPDH* mRNA levels. Non-heat shock control and *AFF4* siRNA-treated cells are shown for comparison. Expression levels were measured by quantitative RT-PCR and normalized to 18S rRNA. Error bars represent standard deviations.

3.2.6. AFF4 is required for the proper expression of MLL chimera target genes.

To begin to investigate the role of AFF4 as a common component of complexes formed by MLL chimeras, I assessed the recruitment of AFF4 to *HOXA9* and *HOXA10* loci in the MV4-11 cell line from a patient with a MLL-AFF1 translocation. As with many MLL translocations, *HOXA9* and *HOXA10* are up-regulated in these cells (Guenther et al., 2008). Indeed, chromatin immunoprecipitation with antibodies corresponding to the C-

terminal portion of AFF1, which is contained in this MLL chimera, shows recruitment to *HOXA9* in the MV4-11 cells (Figure 3.9A), as well as in another MLL-AFF1-rearranged leukemia cell line, SEM, but not in an unrelated leukemia cell line, REH (Figure 3.9B). Interestingly, AFF4 is also recruited to *HOXA9* and *HOXA10* in the MV4-11 and SEM cells, despite the fact that it is the related AFF1 gene that is involved in the MLL translocation in these cells (Figure 3.9A and 3.9B). The antibody to AFF4 was raised against an amino-terminal portion not found in MLL chimeras, ruling out cross-reaction with the related AFF1 protein that is part of the MLL chimera. Besides AFF4, other components of SEC, such as CDK9 and ELL2, are also found at *HOXA9* and *HOXA10* loci in SEM, but not REH cells (Figure 3.9C and 3.9D).

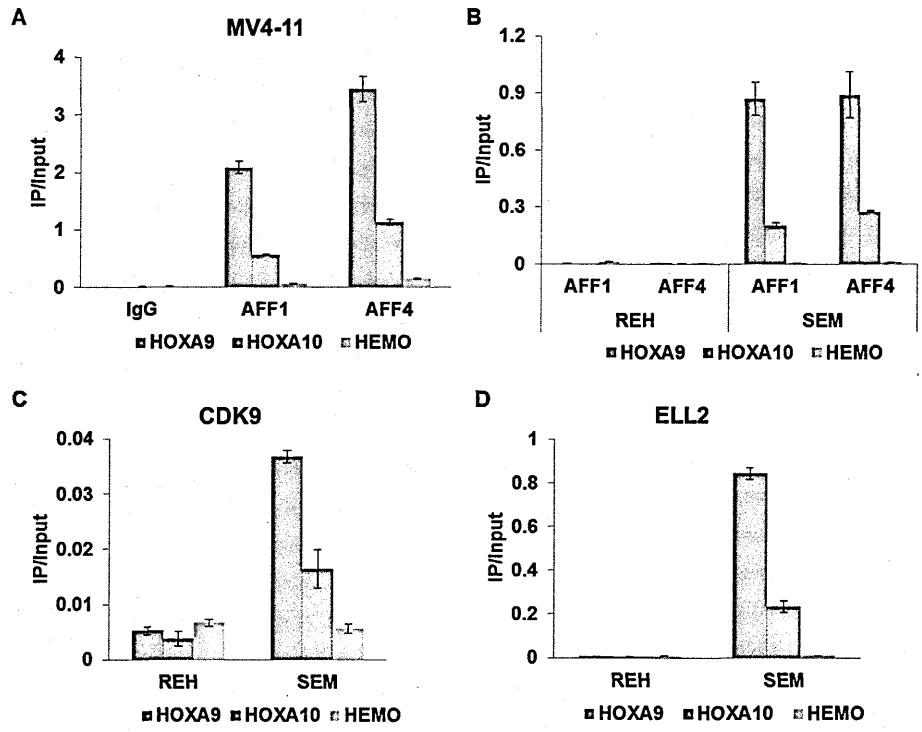


Figure 3.9. AFF4/SEC is recruited to MLL chimera target genes in leukemic cells.
Recruitment of SEC to genes induced by the MLL-AFF1 chimera in MV4-11 and SEM cells, but not the control cell line REH bearing wild-type MLL. (A-B) Antibodies to the C-terminal domain of AFF1 found in the MLL chimera and antibodies raised against the N-terminal domain of AFF4 were used in ChIP-qPCR assays at *HOXA9* and *HOXA10* loci, known targets of the MLL chimera found in human leukemia. As expected AFF1, shows recruitment to *HOXA9* and *HOXA10* in MV4-11 and SEM. AFF4 is also recruited to *HOXA9* and *HOXA10*, consistent with its co-purification with the MLL-AFF1 chimera in Figure 1A and Table 1. (C-D) Other components of SEC, including CDK9 and ELL2, are also found at HOXA9 and HOXA10 genes in MLL-rearranged SEM cells. The beta globin gene (HEMO), which is not expressed in MV4-11, SEM, and REH cells, is used as a negative control in (A-D).

To assess the functional significance of AFF4 recruitment to MLL-AFF1 target genes, I performed lentiviral delivery of AFF4 shRNA to the MV4-11 cells (Figure 3.10A). Significant reductions of *HOXA9* and *HOXA10* are observed upon knockdown of AFF4 in these cells (Figure 3.10B). These findings lend support to our hypothesis that AFF4, a very rare translocation partner of MLL, is nonetheless a component of many MLL-fusion protein complexes and participates in leukemogenesis.

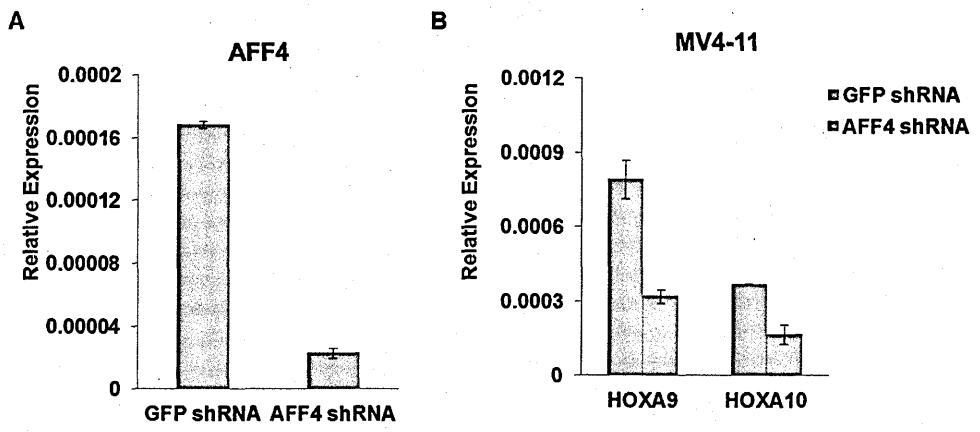


Figure 3.10. AFF4 is required for the proper expression of MLL chimera target genes in leukemic cells.

(A) Knockdown of AFF4 in MV4-11 by retroviral introduction of a shRNA targeting AFF4. (B) Reduction of *HOXA9* and *HOXA10* expression in MV4-11 cells after AFF4 knockdown. GFP shRNA is used as a non-targeting control shRNA. Expression levels were measured by quantitative RT-PCR and normalized to 18S rRNA. Error bars represent standard deviations.

3.2.7. AFF4 is also required for the proper *MYC* gene expression in leukemia cells.

Recent studies have shown that *MYC* plays an important role in the self-renewal of leukemia stem cells and that the anti-leukemic effect of the BRD4 inhibitor, JQ1, are due to the subsequent reduction of *MYC* expression (Dawson et al., 2011; Delmore et al., 2011; Zuber et al., 2011). However, this effect was only seen in AML cells, with ALL cells being insensitive to JQ1 treatment (Zuber et al., 2011). It has also been reported that MED26, which is involved in the recruitment of SEC, regulates *MYC* gene expression in 293 cells (Takahashi et al., 2011). To investigate the requirement of AFF4 in *MYC* gene expression in leukemic cell lines, we performed AFF4 ChIP in both AML and ALL cell lines. Interestingly, as shown in Figure 3.11A, AFF4 localizes at the *MYC* locus in the cell lines tested. Knockdown of AFF4 further indicates that AFF4 is required for *MYC* gene expression in leukemic cells (Figure 3.11B and 3.11C). Taken together, the above results suggest that as a direct upstream regulator of the *MYC* gene, AFF4 could be a novel drug target for leukemia, functioning in a broader spectrum than BRD4 inhibitors.

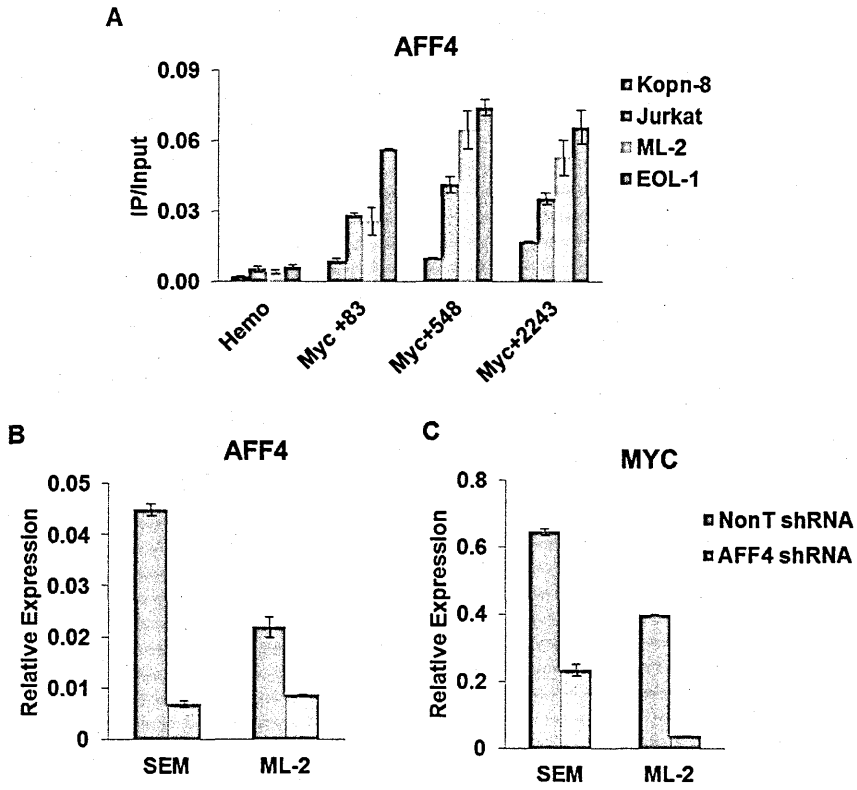


Figure 3.11. *MYC* expression in leukemia cells is regulated by AFF4.
(A) AFF4 is recruited to the *MYC* gene in different leukemia cell lines. ChIP of AFF4 in ALL cell lines (Jurkat and Kopn-8) and AML cell lines (EOL-1 and ML-2) demonstrates the recruitment of AFF4 to the MYC gene in these cell lines. (B) RT-qPCR analysis showing the efficiency of AFF4 knockdown in leukemia cell lines SEM (human ALL with MLL-AFF1 translocation) and ML-2 (human AML cells). (C) RT-qPCR analysis of *MYC* mRNA levels upon AFF4 knockdown in leukemic cell lines. SEM and ML-2 were transfected with NonT or AFF4 shRNA. 72 hours after transfection and puromycin selection, total RNA was extracted and *MYC* mRNA levels were assessed by RT-qPCR. Expression is relative to *GAPDH*. Error bars represent standard deviations. This figure was done by the collaboration with Dr. Zhuojuan Luo.

3.3.Discussion

Previous studies have provided evidence for links among different MLL translocation partners. ENL, AF9, and AF10 have been linked to the histone methyltransferase Dot1; and it was suggested that a common mechanism of MLL

translocation-based leukemia was through H3K79 methylation by Dot1 (Bitoun et al., 2007; Krivtsov et al., 2008; Mueller et al., 2007; Mueller et al., 2009; Okada et al., 2005). However, the most common translocation partner of MLL is AFF1, which my present studies show does not associate with Dot1. Other studies suggesting a physical interaction between Dot1 and AFF1 were based on the isolation of these two proteins in ENL immunoprecipitates and through building a network of 2-hybrid and other interactions. My studies, (Lin et al., 2010) together with our recent purification of the DOT1L-containing complex (Mohan et al., 2010a), indicate that ENL participates in two distinct complexes, one with Dot1 and one within the SEC. MLL-AFF1 does not physically associate with Dot1, so a role for Dot1 at genes up-regulated in MLL-AFF1 leukemias may be subsequent to gene activation by this MLL chimera. Our lab has recently shown that Dot1-mediated H3K79 methylation is linked to cell cycle control in yeast (Schulze et al., 2009) and that methylation by Dot1 could also have some roles in transcriptional enhancement in leukemogenesis. In contrast, MLL-AFF1 co-purified the SEC complex containing ELL1 and P-TEFb, two proven transcription elongation factors *in vitro* and *in vivo*, each with demonstrated abilities to activate transcription through transcription elongation. An important area for future investigation is to define the relative contributions of these two types of complexes to leukemogenesis.

Of the most common MLL partners in leukemia, AFF1, AF9, ENL, and ELL1 were purified as part of the SEC, as was AFF4, a rare MLL partner (Lin et al., 2010). The SEC also contains three other well-known elongation factors: ELL2, ELL3, and the positive transcription elongation factor b (P-TEFb); and the ELL-associated factors, EAF1 and EAF2 (Lin et al., 2010). The three ELL family proteins share a conserved C-terminal occludin domain, which is required for the MLL–ELL chimeras to induce leukemia (DiMartino et al., 2000). The co-purification of ELL1, ELL2, ELL3, AFF1, AFF4, and P-

TEFb from the SEC and the evidence linking them to the control of the transcription elongation checkpoint (TEC) suggests that MLL chimeras (like MLL–AFF1) activate MLL target gene expression through the SEC, perhaps by bypassing normal TEC steps (Lin et al., 2010; Mohan et al., 2010b; Smith et al., 2011a) (Figure 3.12). Indeed, knockdown of the central SEC component AFF4 in *MLL-AFF1* leukemia cells causes a reduction in the expression of *HOXA9*, a key mediator of leukemogenesis (Lin et al., 2010).

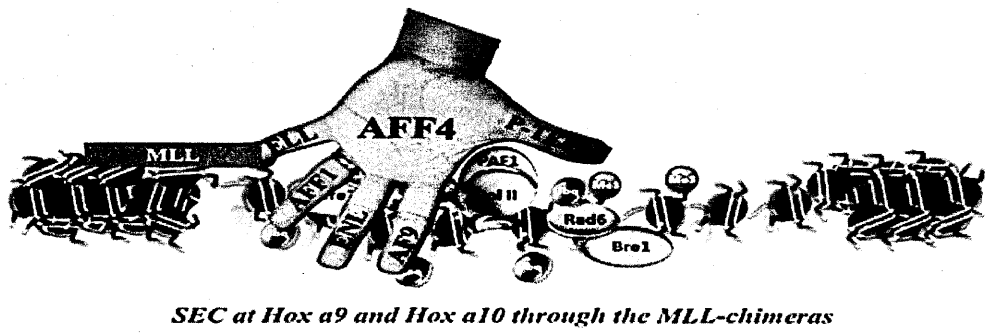


Figure 3.12. Model for SEC in MLL-rearranged leukemia.
In normal hematopoiesis, genes such as *HOXA9* and *HOXA10* are under strict transcriptional control. Translocation of the SEC subunits, such as ENL, ELL, and AFF1 to the MLL N-terminus (MLLn) localizes SEC to these genes and also stabilizes MLL-SEC locally, leading to release of the paused Pol II on these genes without appropriate checkpoints.

One of the direct target genes of SEC identified in the present study is *MYC*, which is a master regulator of cell cycle and proliferation, and is overexpressed in many human cancers, thereby implicating SEC in the control of cell proliferation (Eilers and Eisenman, 2008; Meyer and Penn, 2008). *MYC* is one of a few well-characterized genes which are mainly regulated at the level of transcription elongation by promoter- proximal paused Pol II (Krumm et al., 1993; Spencer and Groudine, 1990). Recent findings have indicated that BRD4 is also a critical player in the maintenance of AML through regulating the expression of the *MYC* gene. Knockdown of BRD4 by a specific shRNA or

pharmacological inhibition of the BET bromodomain by the small molecule JQ1 leads to selective suppression of the *MYC*-regulated transcriptional network, prompting cell cycle arrest and apoptosis in AML cell lines, but not ALL cell lines (Dawson et al., 2011; Delmore et al., 2011; Zuber et al., 2011). The proven association of SEC components with leukemogenesis, together with our finding that SEC is required for the expression of *MYC* in both AML and ALL leukemia cells, suggests that AFF4/SEC is a potential therapeutic target for the treatment of leukemia or other types of cancers associated with *MYC* overexpression. However, it needs to be further investigated that if the depletion of AFF4 or SEC will affect the transformation activity of MLL chimera proteins *in vivo*.

Both AFF1 and AFF4 copurify with the ELL proteins and another AFF protein, AFF3, which is also a rare translocation partner with MLL (von Bergh et al., 2002). The related *AFF2* gene (*fragile X mental retardation 2, FMR2*) is silenced in a form of mental retardation (Knight et al., 1993), thus implicating all members of this family in human diseases (Bitoun and Davies, 2009). P-TEFb itself is involved in a number of malignancies and developmental diseases (Romano and Giordano, 2008), and it will be intriguing to determine which of these processes involve SEC or SEC-like complexes (Luo et al., 2012a).

Collectively, the results of this study identify AFF4 as a component of the Pol II elongation complexes consisting of ELLs, P-TEFb, and several of the common MLL fusion partners. These findings could prove critical for understanding the etiology of MLL translocation-based leukemias and for identifying additional targets for the treatment of the hematological malignancies resulting from these translocations, as well as for understanding fundamental aspects of transcription elongation control in development.

Chapter 4. P-TEFb within SEC Regulates Rapid Transcriptional Activation in the Presence or Absence of Paused Pol II

4.1. Introduction

Transcriptional regulation by RNA polymerase II (Pol II) is a multi-faceted process requiring the concerted action of a large number of factors for the steadfast synthesis of full-length messenger RNA (Boettiger and Levine, 2009; Bres et al., 2008; Shilatifard et al., 2003; Sims et al., 2004; Workman and Kingston, 1998). Transcription by Pol II is divided into four stages: initiation, promoter clearance, elongation, and termination. The initiation stage of transcription requires nucleosomal remodeling around the enhancer and promoter regions followed by the recognition of the promoter elements by the basal transcription machinery and Pol II. Once the basal factors and Pol II are recruited to the promoter elements, the catalysis of the first phosphodiester bond marks the initiation of transcription (Shilatifard, 1998; Sims et al., 2004). For many years, it was considered that transcription initiation was the rate-limiting step to the transcription process as a whole. However, a large number of studies demonstrated that the elongation stage of transcription catalyzed by a number of factors is essential for productive transcription (Levine, 2011; Shilatifard et al., 2003; Sims et al., 2004). In support of a vital role for the elongation stage of transcription in development, it has been demonstrated that the perturbation of this stage of transcription or the factors involved in this process results in the pathogenesis of human diseases including cancer (Mohan et al., 2010b; Shilatifard et al., 2003).

In addition to the control of the productive elongation stage of transcription by Pol II elongation factors, many developmentally regulated genes are marked by stalled or paused Pol II at their promoter-proximal regions (Boettiger and Levine, 2009; Core et al.,

2008; Muse et al., 2007; Rahl et al., 2010; Zeitlinger et al., 2007). Such polymerases have already been initiated and are awaiting proper developmental signals to enter the processive stage of transcription elongation (Rougvie and Lis, 1988). Some studies have suggested that marking such developmentally regulated genes by paused Pol II could enhance their ability to be induced rapidly in a robust manner (Nechaev and Adelman, 2008). Other studies, however, have proposed that the presence of paused Pol II at developmentally regulated genes allows for a synchronous induction of the same set of genes in distinct cell populations at the appropriate stage of development (Boettiger and Levine, 2009).

Multiple factors have been identified to achieve proper promoter clearance and the processive elongation stage of transcription during development. These factors include Elongin A, DSIF, NELF, P-TEFb, and ELL (Bres et al., 2008; Jones and Peterlin, 1994; Levine, 2011; Peterlin and Price, 2006; Shilatifard et al., 2003; Sims et al., 2004). Both Elongin A and DSIF are capable of increasing the catalytic rate of the productive transcription by Pol II, however, in addition to its role in this process, DSIF also works with NELF to regulate Pol II arrest (Cheng and Price, 2007; Shilatifard et al., 2003; Yamaguchi et al., 1999). Such arrested Pol IIs are released by the Cdk9 kinase activity of P-TEFb, which phosphorylates the CTD of Pol II and many of the other transcription factors signaling the release of the stalled Pol II into productive transcription (Fuda et al., 2009; Jones and Peterlin, 1994). ELL was purified based on its catalytic properties to increase the V_{max} rate of transcription elongation by Pol II (Shilatifard, 1998; Shilatifard et al., 1996). Translocation of ELL involving the mixed lineage leukemia gene, *MLL*, is associated with the pathogenesis of childhood leukemia and the misregulation of developmental genes (Thirman et al., 1994).

P-TEFb participates in a variety of complexes, both active and inactive (Bres et al., 2008; He and Zhou, 2011). Both Brd4 and Myc-containing P-TEFb complexes have been considered to be major regulators of transcription elongation (Donner et al., 2010; Rahl et al., 2010; Zhou and Yik, 2006; Zippo et al., 2009). To investigate to what degree SEC functions genome-wide in transcription elongation control, I performed ChIP-seq studies in both mouse embryonic stem (ES) cells in response to retinoic acid induction and human HCT-116 cells in response to serum stimulation. My studies in mouse ES cells identified gene targets for SEC, many of which are developmental regulators with paused Pol II that were rapidly induced to high, but relatively uniform, levels. My studies in human HCT-116 cells found that SEC is also a major regulator of immediate early genes induced by growth factors.

Together, these findings suggest that the presence of paused Pol II at promoter-proximal regions and recruitment of SEC upon activation may represent a major cellular mechanism for rapid and uniform induction of gene expression upon exposure to key developmental signals. Intriguingly, my global genomic studies in ES cells also identified a requirement for SEC at *cytochrome P450 26A1* (*Cyp26a1*), a gene which does not bear paused Pol II at its promoter-proximal region, yet responds dynamically to RA in an even more rapid manner than other genes that have paused Pol II at their promoter-proximal regions. My findings suggest that the recruitment of SEC allows genes to respond in a rapid and dynamic manner to developmental signals in different cell types in mammals, and that SEC is involved in transcriptional induction that is both dependent and independent of the presence of paused Pol II.

4.2.Results

4.2.1. Genome-wide occupancy analyses of SEC components

To investigate a possible role of SEC in the control of developmental genes poised for activation in early development, I developed antibodies to the SEC components and performed a genome-wide occupancy analysis of the SEC components in mouse ES cells using ChIP-seq of AFF4, ELL2, CDK9, and RNA Pol II. These SEC components co-occupy many of the same genes, including highly expressed housekeeping genes such as the histone genes (Figure 4.1), however, SEC is only found at a subset of highly transcribed genes (Figures 4.2A-4.2B). SEC components are enriched at the transcription start site (TSS) regions of these genes and within the gene body similar to the Pol II distribution (Figures 4.2C-D). I notice that the distribution pattern of ELL2 is more 5'-end bias compared with the distribution of AFF4. It is possible that ELL2 and AFF4 could function in different steps of transcription but with a coordinated way. The co-occupancy of the AFF4 and ELL2 components of SEC correlates with a high level of expression of genes in mouse embryonic stem cells (Figure 4.2D) suggesting that SEC is frequently associated with highly transcribed regions.

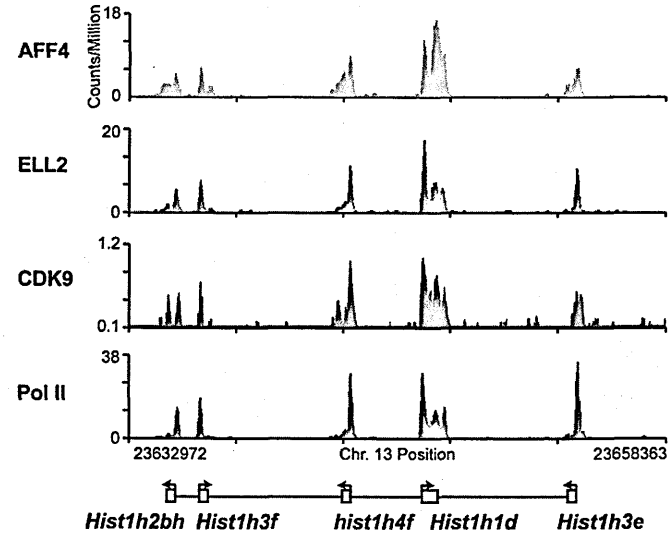


Figure 4.1. SEC enriched at highly transcribed genes such as the histone loci.

SEC subunits are enriched at the transcription start sites (TSS) of these genes, but can also travel with Pol II into the gene body.

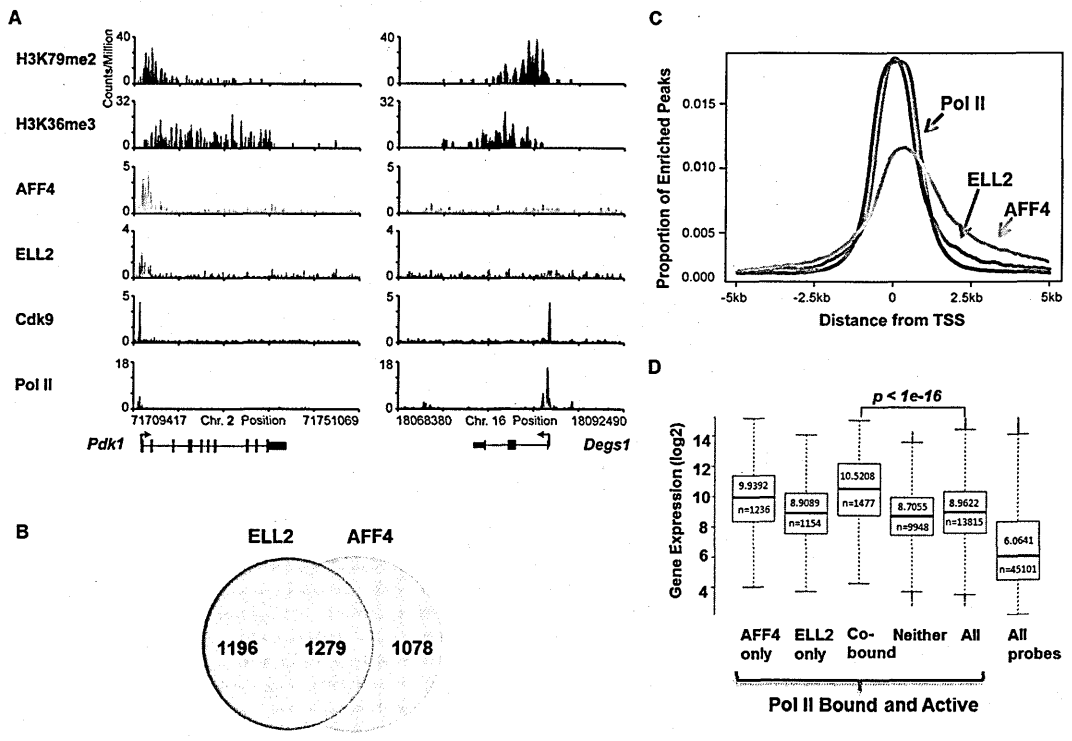


Figure 4.2. Global occupancy of the SEC subunits in mouse embryonic stem cells.
(A) Genome-wide analysis of SEC components AFF4, ELL2, and CDK9 by ChIP-seq in ES cells finds SEC enriched at a subset of actively transcribed genes. Shown are two genes with high levels of expression in ES cells. Left panel: The *Pdk1* gene is occupied by the SEC subunits ELL2 and AFF4. Right panel: The *Degs1* gene does not have significant levels of the SEC components AFF4 and ELL2. H3K36me3 and H3K79me2 data from (Marson et al. 2008) are shown as markers of actively transcribed genes. (B) Venn diagram analysis of AFF4 and ELL2 occupied genes. Around 50% of AFF4-enriched genes are also occupied by ELL2, demonstrating that in mouse ES cells, these two proteins share a similar global occupancy. (C) Histogram of the genome-wide occupancy of AFF4, ELL2, and Pol II. The canonical TSS of each gene in the genome was used to measure the distance to the nearest bound region, which is plotted if falling within 5kb of the TSS. This analysis shows that SEC components are enriched over the TSS, similar to Pol II occupancy. (D) AFF4 and ELL2 co-occupy highly transcribed genes. The dark lines in the boxplots, and the number above the line, indicate the median level of expression for the gene subset indicated. The number below the line indicates the number of Affymetrix probe sets that correspond to the gene subset. Probe sets for ELL2 and AFF4 co-bound genes show significantly higher expression compared to all Pol II-bound and active genes ($p<1e-16$ by Wilcoxon two-sample rank sum test). The gene subset containing neither AFF4 nor ELL2 also shows some highly expressed genes. Genes were called active if they were determined present on the array by the MAS5 algorithm. The genome-wide data was done by the collaboration with Alexander Garrus.

4.2.2. SEC is recruited to paused *Hoxa* cluster genes upon RA treatment in ES cells.

The SEC was discovered based on the purification of several of the MLL chimeras that are commonly found in MLL-rearranged leukemias (Lin et al., 2010). It is not currently known why SEC components are so frequently found to be translocation partners with MLL. One possibility is that the downstream target genes misregulated by MLL chimeras, such as the *HOX* genes, are also regulated by SEC during normal hematopoietic development. In MLL translocated leukemias, these genes become misregulated due to the inappropriate recruitment or misregulation of SEC by MLL chimeras, resulting in the premature activation of transcription by releasing paused Pol II (Lin et al., 2010; Mohan et al., 2010b; Smith et al., 2011b). For example, the MLL-AFF1 translocation can relocalize SEC to the *HOXA9* and *HOXA10* genes. Many developmentally regulated genes in flies and mammals have paused Pol II at the TSS before their activation during development (Muse et al., 2007; Zeitlinger et al., 2007). In mammalian stem cells, these genes are characterized by a bivalent mark of both H3K4 and H3K27 trimethylation on the same gene (Bernstein et al., 2006). Looking within the *Hox* clusters in ES cells, we find bivalent marks co-occurring with Pol II and the general transcription factors at the TSS at four of the *Hoxa* cluster genes, *Hoxa1*, *Hoxa3*, *Hoxa4,* and *Hoxa7*, but not at the promoters of the *Hoxb* genes (Figure 4.3).

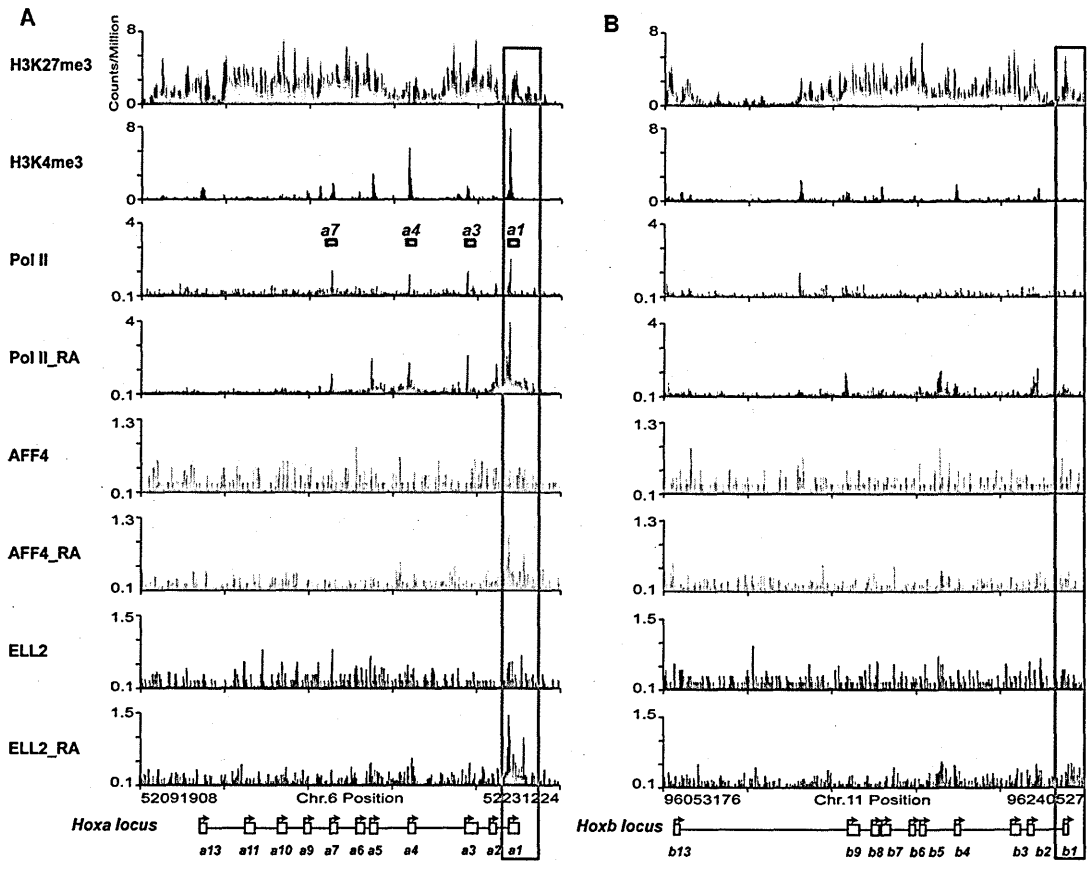


Figure 4.3. The *Hoxa1* promoter is preloaded with Pol II and recruits SEC after RA treatment in ES cells.

(A) Bivalent marks, paused Pol II and SEC recruitment to the *Hoxa* cluster. In ES cells, the whole *Hoxa* cluster is highly enriched for H3K27me3, and also contains H3K4me3 at the promoters of a subset of genes, including *Hoxa1*, *a3*, *a4* and *a7*. These regions are pre-loaded with Pol II (bars indicate regions which have both a bivalent mark and Pol II). (B) Bivalent marks and paused Pol II are both largely absent from the *Hoxb* genes, which do not recruit SEC after 6 hour RA treatment. While H3K27me3 marks the whole cluster of *Hoxb* genes, only *Hoxb4, b7, and b9* contain H3K4me3 at their promoters and can be considered bivalent. There is no significant Pol II detected on the promoters of the *Hoxb* genes in ES cells. The bar marks a peak of significant Pol II that doesn't correspond to a known gene feature. Before RA treatment, there is no detectable AFF4 and ELL2 signal on the *Hoxa* or *Hoxb* cluster genes. Both AFF4 and ELL2 are recruited to the *Hoxa1*, but not the *Hoxb1*, gene promoter after exposure to RA for 6 hours. Blue boxes highlight the *Hoxa1* and *Hoxb1* genes. Expanded views of the *Hoxa1* and *Hoxb1* regions are shown in Figure 4.4.

The regulation of gene transcription at the level of paused Pol II, and its controlled release, has been best studied at the heat shock genes such as *HSP70*, as well as in the control of HIV transcription, and both processes require SEC (He et al., 2010; Lin et al., 2010; Sobhian et al., 2010). Genes with paused Pol II such as *HSP70* are transcriptionally

engaged but paused at the 30-40 nucleotides downstream of the TSS, waiting for the proper signals or environmental cues to trigger their rapid transcriptional activation (Giardina et al., 1992; Gilmour and Lis, 1986; Lis, 1998; Rasmussen and Lis, 1993, 1995). These genes contain basal transcriptional machinery at their promoters, have a form of Pol II phosphorylated on Ser5, but not Ser2, of the CTD, and the Pol II is associated with DSIF/NELF and TFIIB (Nechaev and Adelman, 2011). By all of these criteria, *Hoxa1*, but not *Hoxb1,* is occupied and engaged by paused Pol II (Figure 4.4).

This led me to ask whether SEC was differentially recruited to *Hoxa1* and not *Hoxb1* upon induction by RA treatment. I performed AFF4, ELL2, and Pol II ChIP-sequencing after 6 hours of RA treatment of mouse ES cells. Pol II was recruited to both *Hoxa1* and *Hoxb1* promoters after exposure to RA for 6 hours (Figure 4.3). Interestingly, AFF4 and ELL2 were only recruited to *Hoxa1,* and not *Hoxb1,* by 6 hours of RA treatment (Figure 4.3). However, I cannot rule out the possibility that SEC was not detected at *Hoxb1* due to lower levels of Pol II, and a concomitant decrease in SEC that falls below the detection level. My genome-wide analyses suggest that my ability to detect SEC occupancy on a gene is not strictly dependent on levels of Pol II or transcription levels (Figure 4.2A right panel and Figure 4.2D).

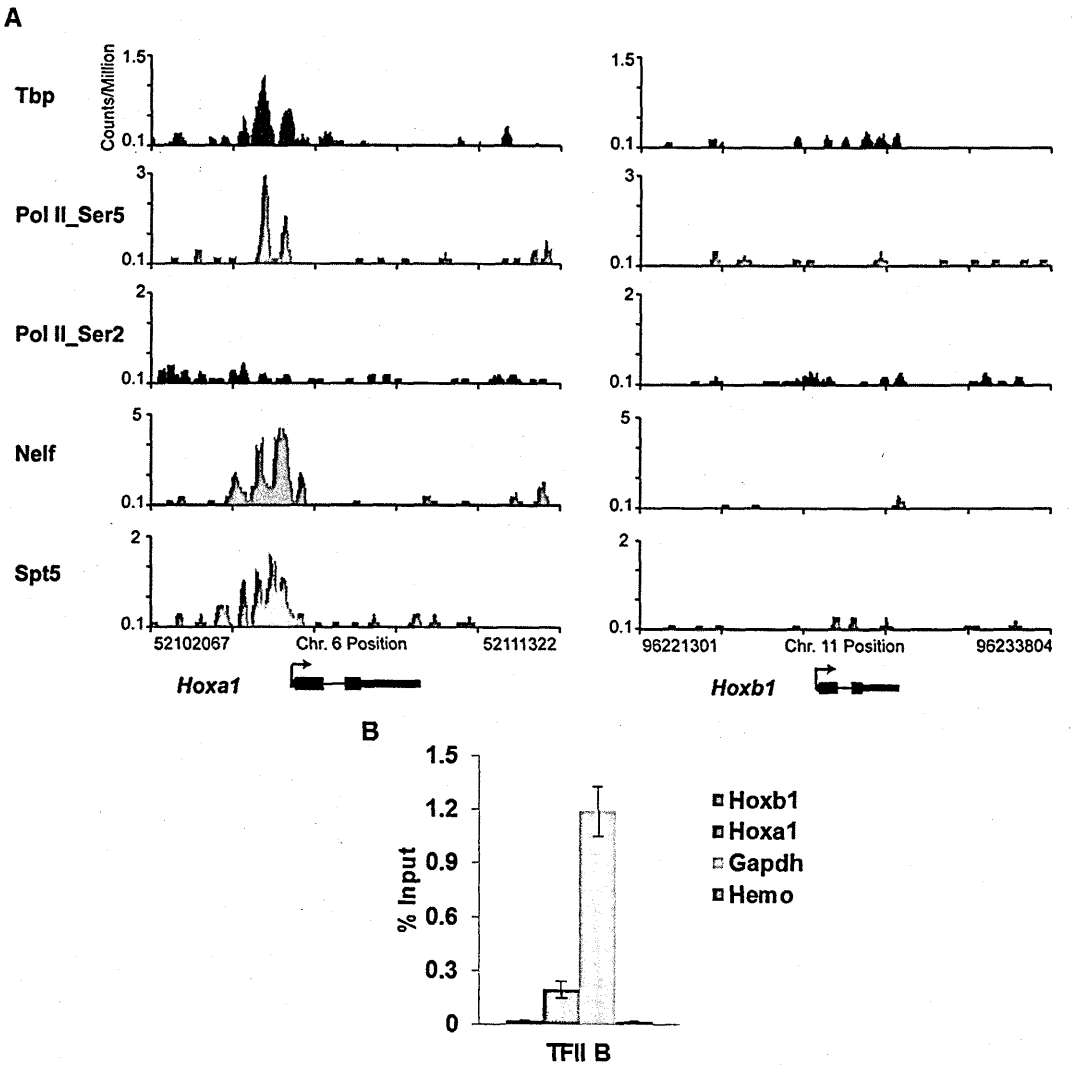


Figure 4.4. *Hoxa1*, but not *Hoxb1*, contains the pausing form of RNA Polymerase II in untreated mouse ES cells.

(A) The *Hoxa1* promoter was preloaded with the S5, but not S2, phosphorylated form of Pol II indicative of TFIIH activity. Also present on *Hoxa1* are DSIF (represented by Spt5) and NELF (represented by NELFA). In contrast, *Hoxb1* is devoid of any of these factors. ChIP-seq data are from (Rahl et al., 2010). (B) The general transcription factor TFIIB is present at the *Hoxa1* before RA treatment, but not at the *Hoxb1* promoter, by ChIP analysis. However, little or no TBP was detectable on the *Hoxa1* and *Hoxb1* gene promoters. *Gapdh* is a highly expressed gene and *Hba1* is a non-transcribed gene in mES cells and these serve as positive and negative controls, respectively. Error bars represent the standard deviation.

4.2.3. SEC is required for the rapid induction of *Hoxa1*.

Promoter proximally paused Pol II has been proposed to allow for a more rapid induction of genes upon differentiation cues (Nechaev and Adelman, 2008). Therefore, we

assayed the induction kinetics of *Hoxa* and *Hoxb* cluster genes by RT-qPCR after RA treatment from 2-24 hours (Figure 4.5A-B). We found that *Hoxa1* and *Hoxb1* were the first genes rapidly induced within their respective clusters, followed more slowly by other members of the clusters in general agreement with the co-linearity of expression that occurs during normal embryonic development (Duboule and Dolle, 1989; Graham et al., 1989; McGinnis and Krumlauf, 1992). The *Hoxa1* and *Hoxb1* paralogs functionally synergize in regulating the hindbrain pattern formation and cranial nerve patterning (Gavalas et al., 2001). During normal mouse development, *Hoxa1* is the first *Hox* gene expressed in neural tissue directly induced by RA through a retinoic acid response element (RARE) located at its 3'-end. It is closely followed by RA-mediated induction of *Hoxb1* through a similar 3' RARE. Hoxa1 also participates in the proper activation of *Hoxb1* by binding to *Hoxb1's* auto regulatory element *(ARE)* located at its 5' region, and Hoxb1 further stimulates transcription of its own gene (Popperl et al., 1995; Studer et al., 1998). Indeed, when looking within the first six hour window of RA treatment of mouse ES cells, we observe that *Hoxa1* is induced more rapidly than *Hoxb1,* mirroring their normal kinetics of induction in neural development (Figure 4.5A-B, blue boxes). The more rapid induction of the *Hoxa1* locus compared with *Hoxb1* could result from the presence of paused Pol II before RA treatment.

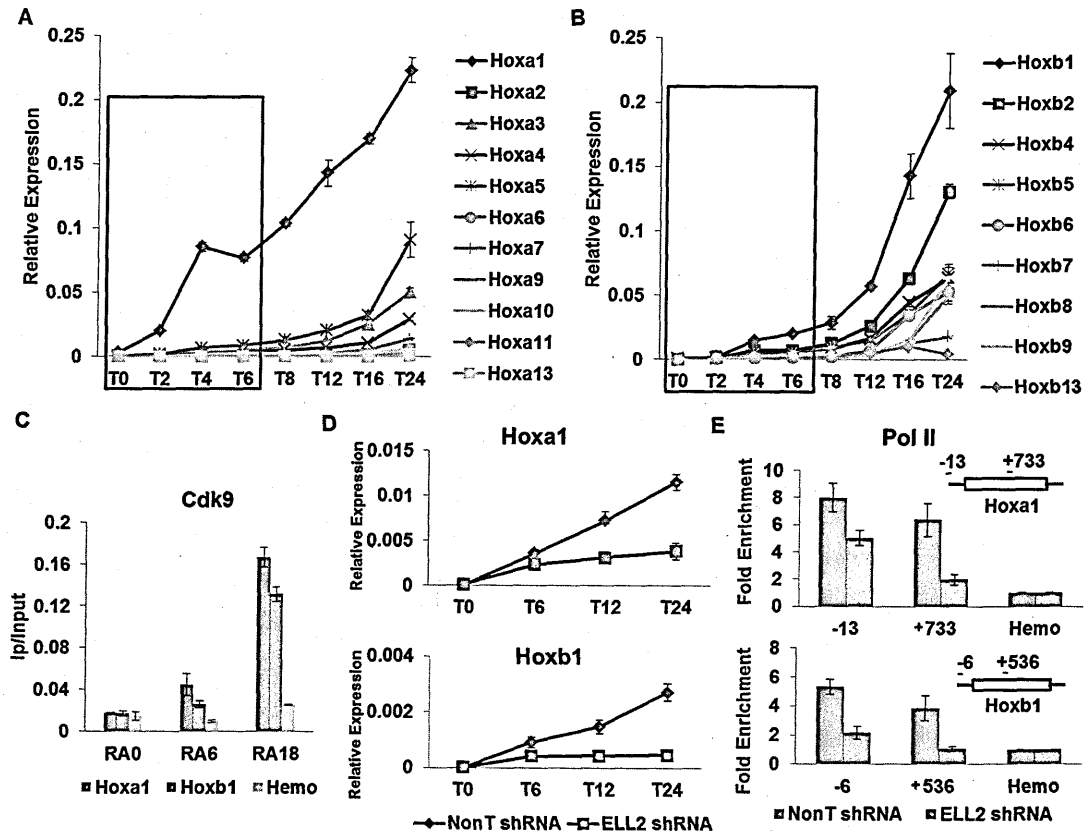


Figure 4.5. SEC is required for the rapid induction of the *Hoxa1* gene.

(A, B) RT-qPCR analysis of *Hoxa* and *Hoxb* cluster genes upon RA treatment. ES cells were treated with RA for different time points as indicated. Total RNAs were extracted from these cells and then subjected to RT-qPCR analysis using an Applied Biosystems' custom Taqman array card. *Hoxa1* was the first *Hox* gene to be induced by RA. Compared with *Hoxa1*, the induction of *Hoxb1* was much slower within the first 6 hours of RA treatment. The blue boxes indicate the first three RA induction time points. (C) Cdk9 is recruited to both the *Hoxa1* and *Hoxb1* gene promoters. Cdk9 ChIP was performed to measure its enrichment on *Hoxa1* and *Hoxb1* after RA treatment. A hemoglobin gene, *Hba* (Hemo), serves as a non-transcribed control gene. (D) ELL2 RNAi inhibits the induction of *Hoxa1 and Hoxb1* by RA. shRNA targeting ELL2 or non-targeting shRNA (NonT) was introduced by lentiviral infection for 3 days before RA treatment. (E) Knockdown of ELL2 reduces Pol II occupancy at *Hoxa1* and *Hoxb1* after 6 hours RA treatment. Pol II occupancy was assayed by chromatin immunoprecipitation at the start site of transcription and in the open reading frame of *Hoxa1* and *Hoxb1* in RA-induced cells. Pol II is reduced in the ORF of both *Hoxa1* and *Hoxb1*, and *Hoxb1* also shows dramatically reduced levels of Pol II at its promoter after ELL2 RNAi. The *Hoxa1* promoter, but not the *Hoxb1* promoter, has pre-bound Pol II before RA treatment (see Figure 4.3). Error bars represent the standard deviation. The Figures 4.5 A and 4.5B were performed by Bony De Kumar from Robb Krumlauf laboratory.

Since SEC was only detected at the *Hoxa1* promoter, and not at the *Hoxb1* promoter, I asked if the Pol II CTD kinase, Cdk9, was also differentially localized to these two genes at early induction time points. Direct comparisons of Cdk9 at *Hoxa1* and *Hoxb1* show that Cdk9 is recruited to both genes as early as 6 hours and has increased occupancy at 18 hours (Figure 4.5C). When ES cells are induced with RA for various time points, in the presence or absence of the Cdk9 inhibitor flavopiridol (Chao and Price, 2001), the induction of both *Hoxa1* and *Hoxb1* are diminished (Figure 4.6). This indicates that Cdk9 is required for the activation of both genes, even though the kinetics of their induction differ. These results also suggest that the recruitment of P-TEFb within SEC, specifically to *Hoxa1,* functions in its rapid induction. In support of this statement, Ell2 RNAi also reduces the induction of *Hoxa1* (Figure 4.5D). Ell2 is also required for the induction of *Hoxb1;* however, this observation could be explained by the requirement of the Hoxa1 protein for the full induction of *Hoxb1* (Studer et al., 1998).

Accordingly, in the absence of Ell2 (Ell2 RNAi), I also detect the loss of Pol II in the body of the *Hoxa1* gene with no significant change or slight reduction in occupancy of Pol II at the *Hoxa1* promoter (Figure 4.5E, upper panel). Furthermore, since *Hoxb1's* expression requires Hoxa1 activity and lacks prior paused Pol II in ES cells, in the absence of Ell2, I detect a loss in Pol II occupancy both at the promoter and in the body of the *Hoxb1* locus (Figure 4.5E, bottom panel). Therefore, *Hoxa1* is likely to be a direct target of SEC, and *Hoxb1* is likely to be an indirect target of SEC. In summary, given the fact that *Hoxa1*, and not *Hoxb1*, possesses paused Pol II and recruits SEC upon a differentiation signal, I hypothesized that the recruitment of SEC to genes bearing paused Pol II is associated with rapid induction.

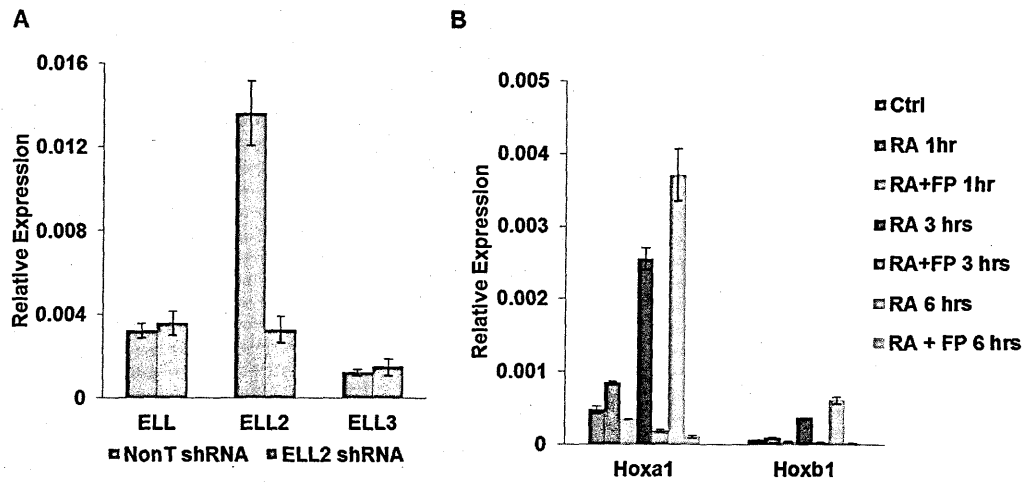


Figure 4.6. CDK9 is required for both *Hoxa1* and *Hoxb1* gene activation by RA.

(A) ELL2 mRNA is specifically and significantly knocked down by ELL2 shRNA. Either shRNA targeting ELL2 or a non-targeting shRNA (NonT) was introduced by lentiviral infection for 3 days. RT-qPCR was used to measure the mRNA levels of *ELL*, *ELL2,* and *ELL3*. (B) The Cdk9 inhibitor, flavopiridol, inhibits the activation of both *Hoxa1* and *Hoxb1* by RA treatment. ES cells were induced with RA for 1, 3, and 6 hours in the presence and absence of 1 μM of flavopiridol. RT-qPCR was used to measure the mRNA levels of *Hoxa1* and *Hoxb1* at the indicated time points. Error bars represent the standard deviation.

4.2.4. SEC is required for the induction of other rapidly induced genes.

Using genome-wide approaches, I asked whether there were other genes that were regulated similarly to *Hoxa1*. I performed gene expression analyses of ES cells treated for 2-6 hours with RA using Affymetrix expression arrays with probes representing *ca.* 30,000 genes (Figure 4.7A). Sorting the gene expression data by fold expression over time showed that only a small number of genes demonstrated rapid and sustained induction over this time frame in a manner similar to *Hoxa1* (Figures 4.7A-B). I found that 37 genes were rapidly induced at least two fold at 2, 4, and 6 hours post induction (Figures 4.7A-B). Among these genes was *Hoxb1*, which our RT-qPCR data had shown was not as rapidly induced as *Hoxa1* (Figures 4.7A-B). I, therefore, performed RT-qPCR analyses of other genes from the top of this list to confirm their patterns of induction (Figure 4.7C). These RT-qPCR studies demonstrated that two of the genes identified by microarray, *Dleu7* and *Csn3* (Figure 4.7C, blue), behaved similarly to *Hoxb1*, and were not as rapidly induced as

Hoxa1, while two others showed the kinetics of rapid induction similar to *Hoxa1* (Figure 4.7C, yellow).

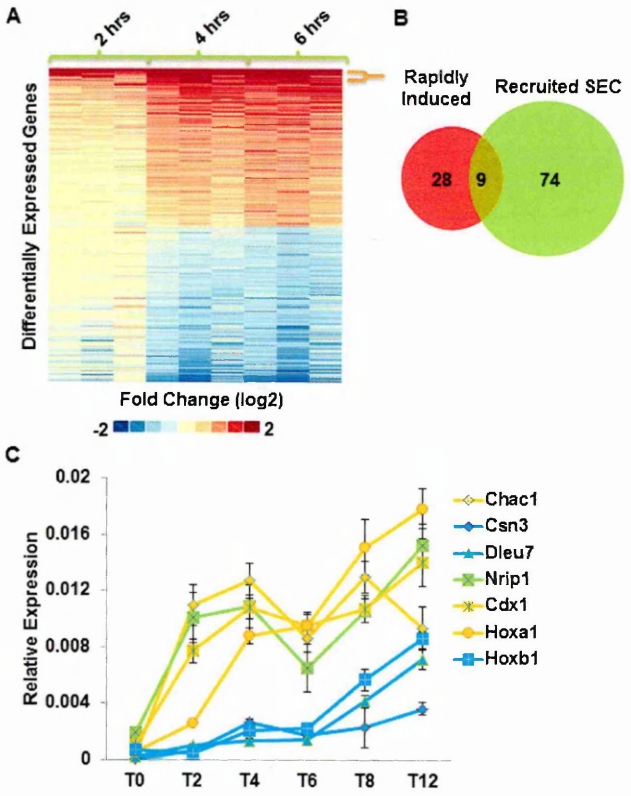


Figure 4.7. SEC regulates the rapid induction of retinoic acid signaling.

(A) Left panel: microarray analyses of RA induction of ES cells as a function of time (2, 4, and 6 hours) in biological triplicate. Differentially-expressed probes (two-fold or more) at the 6th hour post induction compared to no induction are shown. Thirty-seven genes were induced 2 fold or more at each of the 2, 4, and 6 hour time points (demarcated by the orange bracket). (B) Of the 37 induced genes, 9 of them recruited SEC (ELL2 and AFF4). Newly recruited SEC genes are co-bound at 6 hour post induction and not co-bound before induction. (C) RT-qPCR analysis of some of the induced genes identified from the microarray analysis. ES cells were treated with RA for the indicated time points, 0 (T0), 2 (T2), 4 (T4), 6 (T6), 8 (T8), and 12 (T12) hours. Genes that recruit SEC are shown in yellow and genes that do not recruit SEC are shown in blue. *Nrip1*, which doesn't recruit SEC, but is rapidly induced, is shown in green. Error bars represent the standard deviation. The microarray analysis in Figure 4.7A was performed by Bony De Kumar from Robb Krumlauf laboratory.

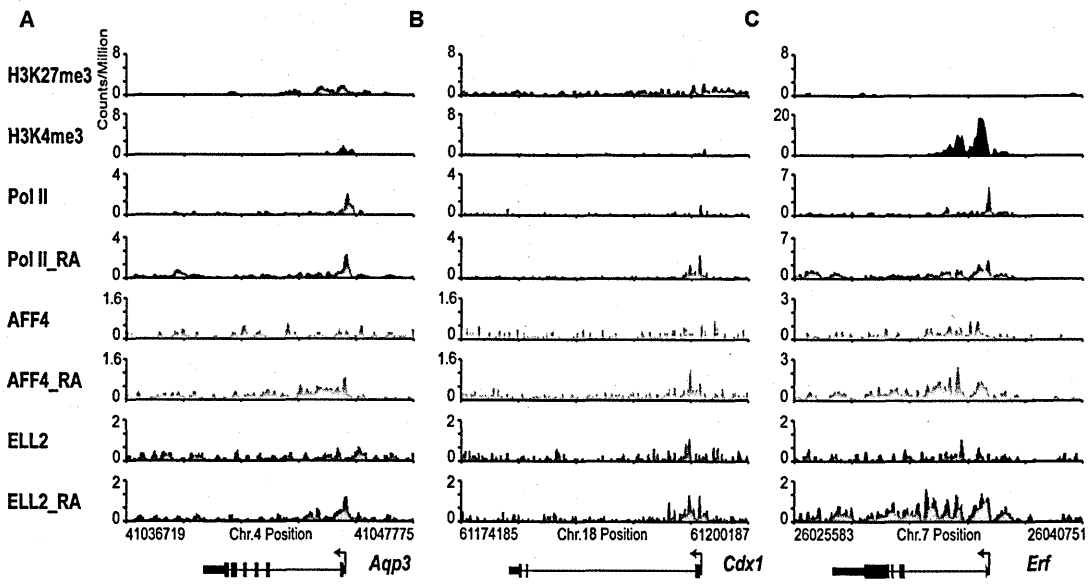


Figure 4.8. SEC is recruited to retinoic acid-induced genes.
(A-C) Examples of ChIP-seq data showing SEC recruitment to RA-induced genes. Shown are *Aqp3*, *Cdx1*, and *Erf*, 3 of the 9 genes from Figure 4.7B.

Many of the induced genes that recruited SEC had bivalent histone marks and paused Pol II prior to RA induction. Genome browser track files for some examples are shown in Figure 4.8A-C. Regardless of whether genes are rapidly or more slowly induced, Cdk9 was recruited and required for their induction (Figure 4.9). This analysis indicates that several genes that recruit SEC with Cdk9 respond more rapidly and uniformly to developmental signals than genes recruiting Cdk9 without SEC. However, the existence of genes like *Nrip1*, which is induced with similar kinetics to *Hoxa1* (Figure 4.7C, green), but does not recruit SEC, suggests that while SEC is a major form of the Cdk9 complexes recruited to genes for rapid gene activation, other pathways to rapid gene activation are also possible.

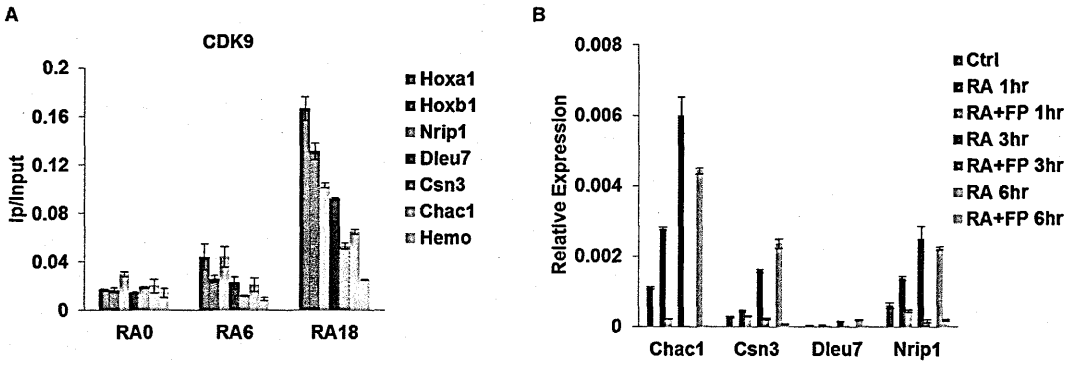


Figure 4.9. The P-TEFb complex is required for all RA highly induced gene activation.

(A) Cdk9 is recruited to all of the RA highly induced gene promoters. Cdk9 ChIP was performed with ES cells in the presence and absence of RA for 6 and 18 hours (RA0, RA6, and RA8, respectively). (B) The Cdk9 inhibitor, flavopiridol (FP), abolished RA-mediated gene activation. ES cells were induced with RA for 1, 3, and 6 hours in the presence and absence of 1 μM of flavopiridol. RT-qPCR was used to measure the mRNA levels at the indicated time points. Error bars represent the standard deviation.

4.2.5. Brd4 is not broadly required for retinoic acid induction of genes.

I tested for the presence of another P-TEFb interactor, Brd4, on these genes and demonstrated that although Brd4 is recruited to these loci upon RA induction (Figure 4.10A), its reduction by RNAi has very little to do with their activation by RA, except for the *Aqp3* gene (Figure 4.10B-C). This observation suggests that although Brd4 is also recruited to those SEC target sites, it might not play a major role for their activation (Figure 4.10A-C). Perhaps, as in the case of HIV-1 transcriptional regulation, Brd4 has a role in maintaining basal levels of transcription, but not in the activation of these genes (Yang et al., 2005).

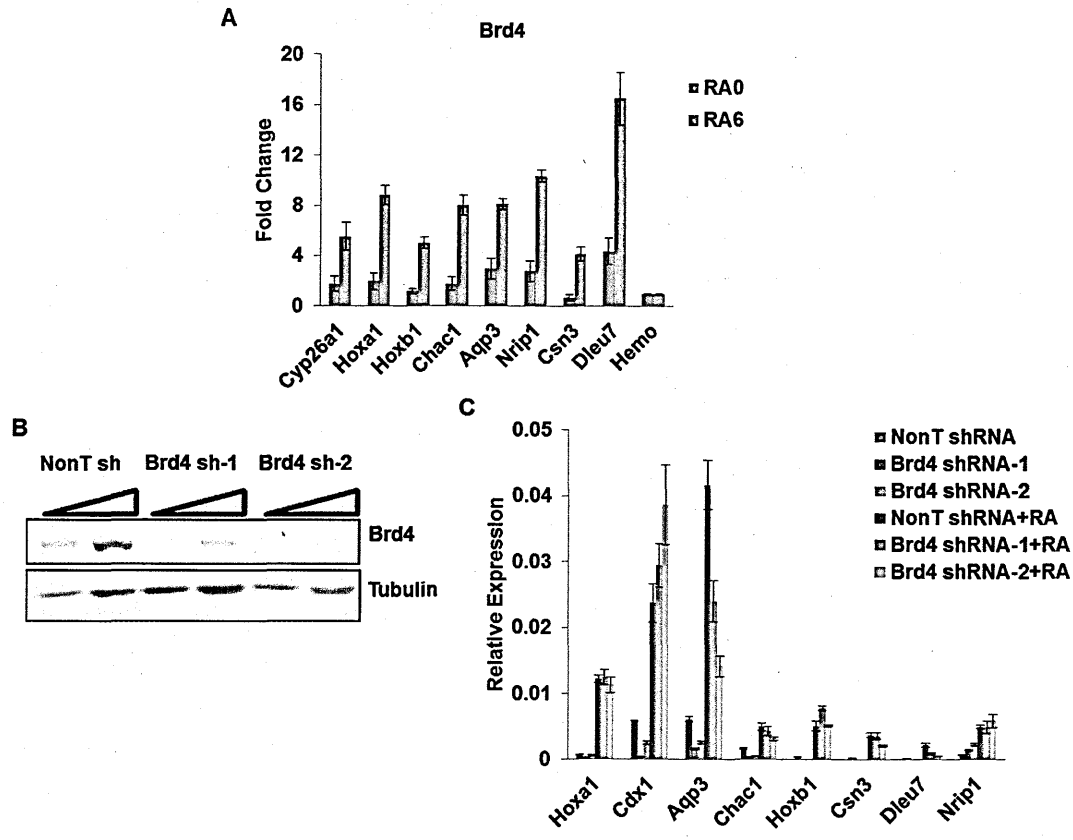


Figure 4.10. Brd4 is broadly present, but not broadly required, for retinoic acid induction of genes.

(A) Chromatin immunoprecipitation of Brd4 at RA-6-induced genes. Brd4 levels significantly increase at all RA-6-induced genes tested. The *Hba* gene serves as a non-transcribed control gene. (B) shRNA-mediated knockdown of Brd4. Two different shRNA constructs targeting Brd4 and a non-targeting shRNA (NonT) were introduced by lentiviral infection for 3 days before RA treatment. Brd4 levels were significantly reduced by Western analysis. Triangles indicate titrations of cell extracts. Tubulin serves as a loading control. (C) Induction of genes with RA is not broadly affected by Brd4 knockdown. Several genes were assayed for expression levels before and after RA treatment. Only *Aqp3* showed a significant decrease in its induction. Error bars represent the standard deviation

4.2.6. SEC is required for the rapid induction of immediate early genes (IEG).

Given the small number of RA-induced genes in the mouse ES system, I sought another system to determine to what degree SEC regulates rapid transcriptional responses to environmental signals. Therefore, I investigated the role of SEC in the induction of genes in response to serum in human cells (Figure 4.11A-C and Figure 4.12). The

immediate early genes (IEG) induced by growth factors are some of the best-characterized genes regulated at the level of the release of paused Pol II (Kong et al., 2005; Simone et al., 2001). I performed ChIP-seq of SEC and Pol II in HCT-116 cells before and after serum stimulation. SEC components are also enriched at TSS in HCT-116 cells, consistent with their distribution in ES cells (Figure 42C and Figure 4.12A). SEC was newly recruited to 55 genes within 30 minutes of serum stimulation (Figure 4.12B). Similarly to what I observed in ES cells (Figure 4.2D), genes bound by AFF4 and ELL2 showed higher levels of expression than those that lacked SEC (Figure 4.12C-D).

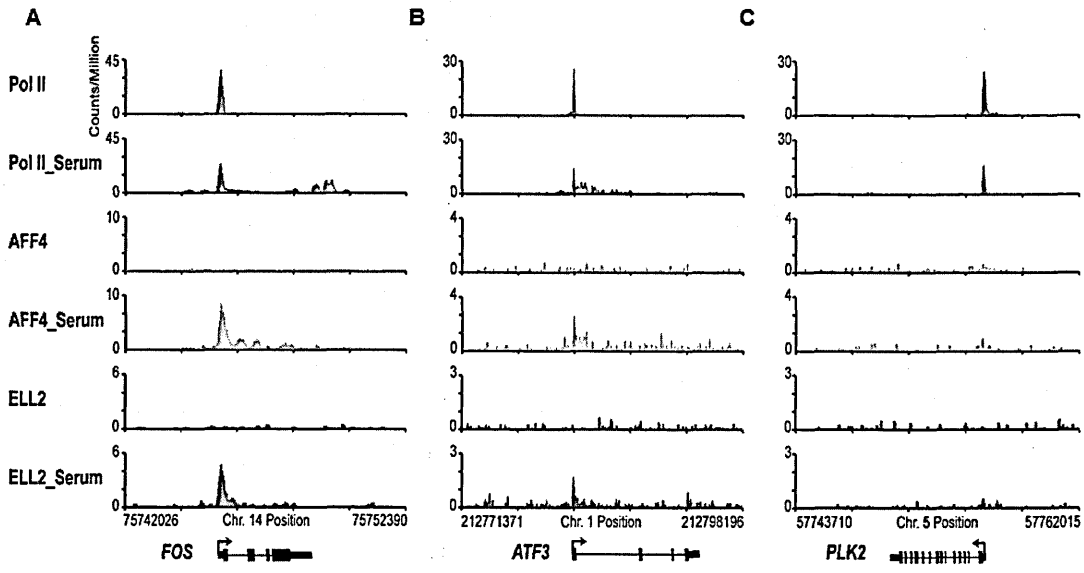


Figure 4.11. SEC is recruited to serum-induced genes.

(A-C) SEC is recruited to the immediate early genes in HCT-116 cells after serum stimulation, genes previously identified as regulated by Brd4-containing P-TEFb complexes.

Previous gene expression analysis of serum inducible genes in HCT-116 cells identified 29 genes that were up-regulated 2 fold or more within 30 minutes of serum stimulation (Donner et al., 2010), 12 of which recruited both AFF4 and ELL2. I also performed RNA-seq analysis in these cells in the presence and absence of serum stimulation and identified 66 genes, which were induced above 2-fold, including 26 out of

the 29 genes identified by Donner et al. To more precisely characterize the induction kinetics of these genes, I performed RT-qPCR on 17 serum-induced genes at different times after serum stimulation (Figure 4.12E). As I had seen with RA induction, serum responsive genes were induced at varying rates, with SEC recruitment frequently occurring on the most rapidly induced genes (Figure 4.12D-E). Thus, SEC appears to be one of the major factors in the rapid release of paused Pol II in response to developmental and environmental stimuli.

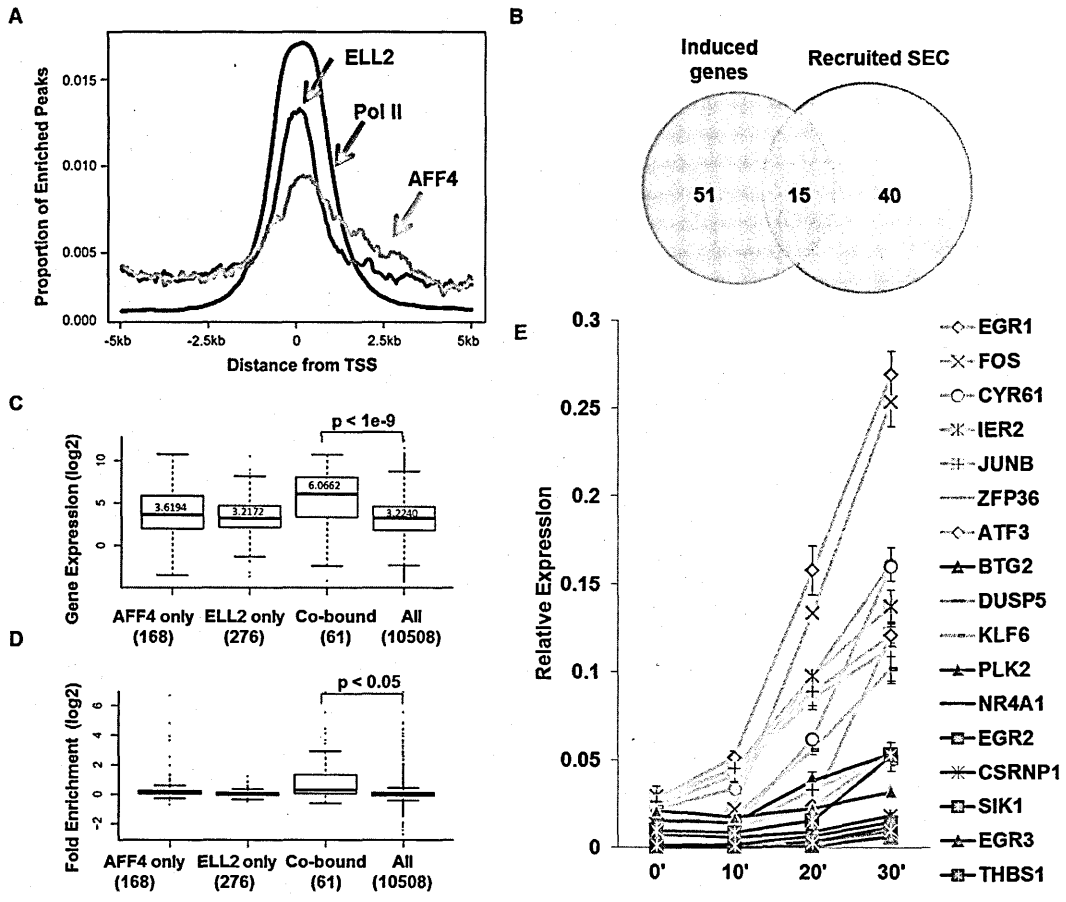


Figure 4.12. SEC is recruited to rapidly induced immediate early genes (IEGs).

ChIP-seq of SEC subunits and Pol II in HCT-116 cells was performed before and after serum stimulation. (A) Histogram of the occupancy of AFF4, ELL2, and Pol II genome-wide. The TSS of each gene in the genome was used to measure the distance to the nearest bound region, which is plotted if falling within 5kb of the TSS. (B) Venn Diagram analysis shows that 15 of the serum-induced genes recruit SEC (AFF4 with ELL2). SEC is newly recruited to 55 genes, where both AFF4 and ELL2 are co-bound after serum stimulation and not co-bound before stimulation. Out of these 55 genes, 15 of them were induced more than 2-fold after serum treatment by RNA-seq analysis. The gene numbers reflect all genes of the above criteria, which were not annotated with the biotype 'pseudogene' or 'processed_transcript'. *(C)* Comparison of RNA-seq expression levels after serum stimulation for gene subsets of all Pol II-bound and active genes. Genes co-bound by SEC show a statistically significant difference in expression versus all Pol II-bound and active genes ($p < 1e-9$ by Wilcoxon rank sum test). Expression is measured as fragments per kilobase of transcript per million reads aligned (FPKM) and shown as the log2 (FPKM). Active genes are defined as having an FPKM >= 0.05. (D) RNA-seq analysis of fold-change of expression after serum stimulation compared to before stimulation for gene subsets of all Pol II-bound and active genes. SEC co-bound genes after serum stimulation show a statistically significant difference in fold-change compared to all Pol II-bound and active genes ($p < 0.05$ by Welch's two sample t-test). (E) RT-qPCR analysis of the induction kinetics of 17 serum-inducible genes. Genes that recruit SEC are shown in yellow and genes that do not recruit SEC are shown in blue. Thus, SEC is frequently associated with the most rapidly activated genes after serum stimulation.

4.2.7. Dynamic and rapid transcriptional induction requires SEC without the presence of paused Pol II.

To date, published studies indicate that paused Pol II functions in the rapid and robust induction of many developmentally regulated genes (Boettiger and Levine, 2009; Muse et al., 2007; Nechaev and Adelman, 2008; Rougvie and Lis, 1988; Zeitlinger et al., 2007). However, my genome-wide expression and ChIP-seq data identified one gene that is extremely rapidly induced by RA, the *Cyp26a1* gene (Figure 4.13). *Cyp26a1* encodes a cytochrome P450 that metabolizes retinoic acid (Duester, 2008). The *Cyp26a1* gene bears several RAREs in its promoter and it is known to be one of the most rapidly induced genes after exposure to RA (Alexander et al., 2009). Loss of *Cyp26a1* is toxic to development in mice, but this toxicity can be rescued by the loss of RA receptor gamma (RARγ) (Abu-Abed et al., 2001; Sakai et al., 2001). While the *Cyp26a1* gene appears to have high levels of H3K27 trimethylation, it contains very low levels of H3K4 trimethylation compared to *Hoxa1* (please see Figure 4.13A and Figure 4.2A-B). Also, this gene lacks paused Pol II in the untreated ES cells (Figure 4.13A). After RA addition, Pol II and SEC are recruited to *Cyp26a1* by 6 hours post induction (Figure 4.13A).

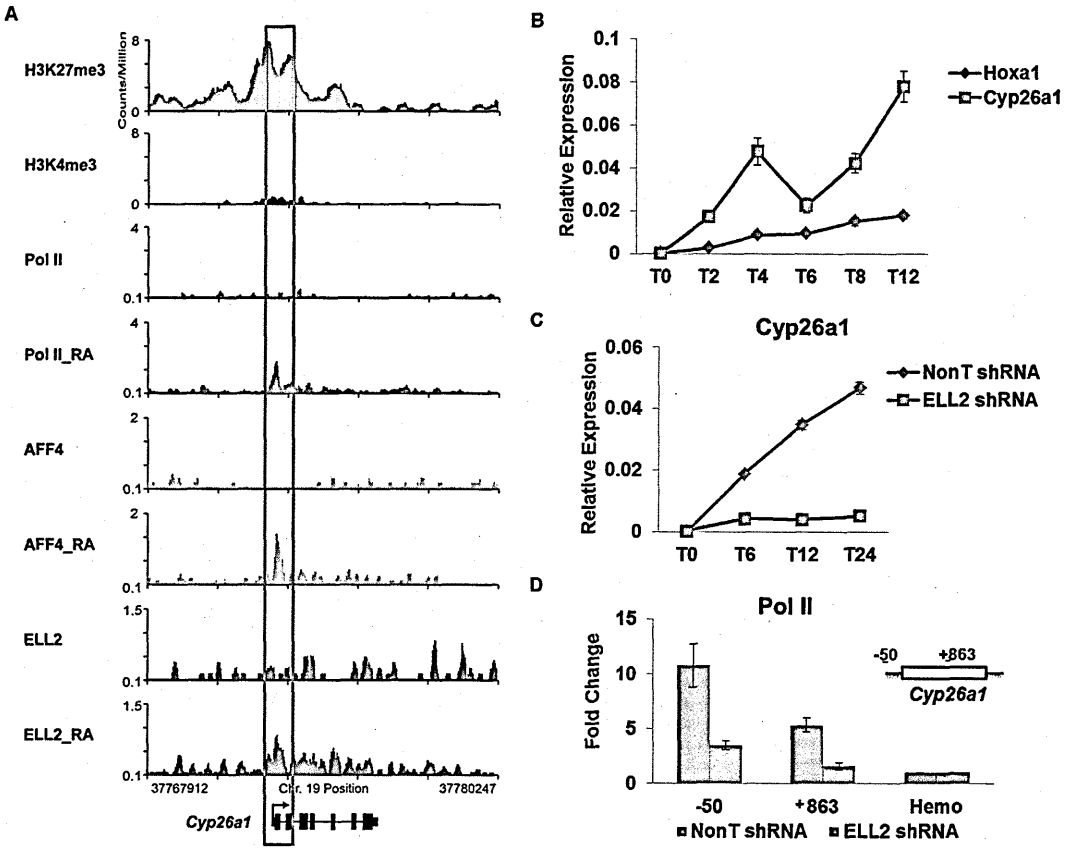


Figure 4.13. The rapid induction of *Cyp26a1* does not involve pre-loaded Pol II.
(A) Pol II, H3K4me3, and H3K27me3 occupancy analysis of the *Cyp26a1* gene before RA induction. Before RA treatment, the *Cyp26a1* promoter is significantly enriched for H3K27me3 with lower levels of H3K4me3. However, there is no detectable Pol II on the promoter. AFF4, ELL2 and Pol II are newly recruited to the *Cyp26a1* gene promoter upon RA treatment. *(B)* RT-qPCR analysis of *Cyp26a1* mRNA levels upon RA treatment. ES cells were treated with RA for the indicated time points, 0 (T0), 2 (T2), 4 (T4), 6 (T6), 8 (T8), and 12 (T12) hours. Total RNAs were extracted from these treated cell samples and then subjected to RT-qPCR analysis. *(C)* ELL2 RNAi inhibits the induction of *Cyp26a1* by RA. shRNA targeting ELL2 or a non-targeting shRNA (NonT) was introduced by lentiviral infection for 3 days before RA treatment. *(D)* Knockdown of ELL2 reduces Pol II occupancy at *Cyp26a1* after 24 hours RA treatment. The *Hba* gene serves as a non-transcribed control gene. Error bars represent the standard deviation.

In mouse ES cells, *Cyp26a1* is more rapidly induced when compared with *Hoxa1* and *Hoxb1* (Figure 4.13B; Figure 4.5C). Knockdown of Ell2 by shRNA treatment causes a reduction in *Cyp26a1* activation and also affects the recruitment of Pol II in its promoter and gene body (Figure 4.13C-D), while flavopiridol completely eliminates *Cyp26a1* induction, indicating that this gene requires Cdk9 for its rapid induction by RA treatment

(Figure 4.14). Furthermore, reduction of the Brd4 level by RNAi did not significantly affect *Cyp26a1* induction, suggesting that it is the SEC version of P-TEFb that regulates this gene. The dynamic induction of *Cyp26a1* without preexisting paused Pol II suggests that there are other mechanisms for rapid induction of transcription during early development, which involves SEC.

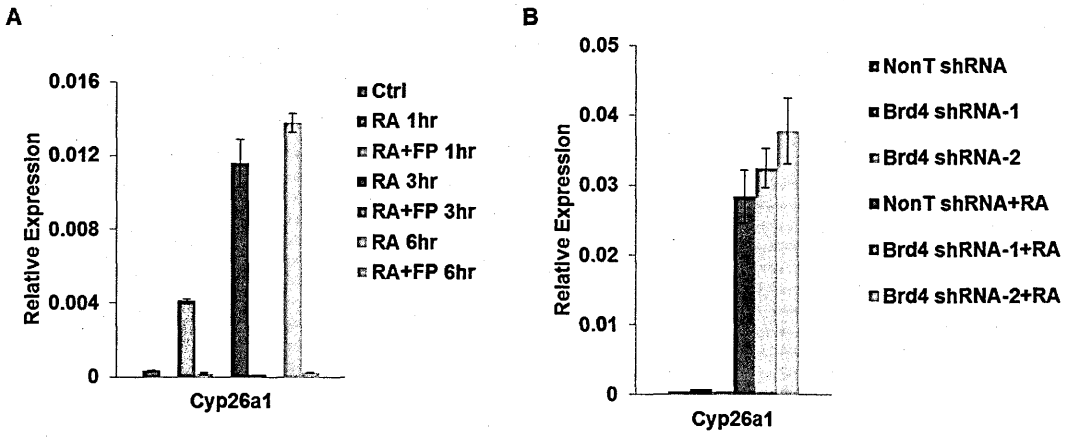


Figure 4.14. The SEC/P-TEFb complex, but not Brd4/P-TEFb complex, is required for the *Cyp26a1* gene activation by RA.

(A) The P-TEFb inhibitor, flavopiridol (FP), abolishes the induction of *Cyp26a1* by RA. ES cells were induced with RA for 1, 3, and 6 hours in the presence and absence of 1 μM of flavopiridol. RT-qPCR was used to measure the *Cyp26a1* mRNA levels at the indicated time points. (B) Induction of *Cyp26a1* with RA is not affected by Brd4 knockdown. Two different shRNA constructs targeting Brd4 and a non-targeting shRNA were introduced by lentiviral infection for 3 days before RA treatment. Total RNAs were extracted from these treated cell samples and then subjected to RT-qPCR analysis. Error bars represent the standard deviation.

4.3. Discussion

Our genome-wide analyses of RA-induced gene transcription and SEC recruitment have identified three classes of genes, two of which require SEC for induction (Figure 4.15). One class, which includes *Hoxb1*, lacks paused Pol II and does not recruit SEC upon induction (Figure 4.15A). A second class, which includes *Hoxa1*, contains paused Pol II, recruits SEC, and is induced more rapidly than the first class (Figure 4.15B). A

third class, exemplified by *Cyp26a1*, recruits SEC, is induced just as rapidly as the second class, but to a greater extent than *Hoxa1*, yet lacks paused Pol II at its promoter-proximal region before induction and requires SEC (Figure 4.15C).

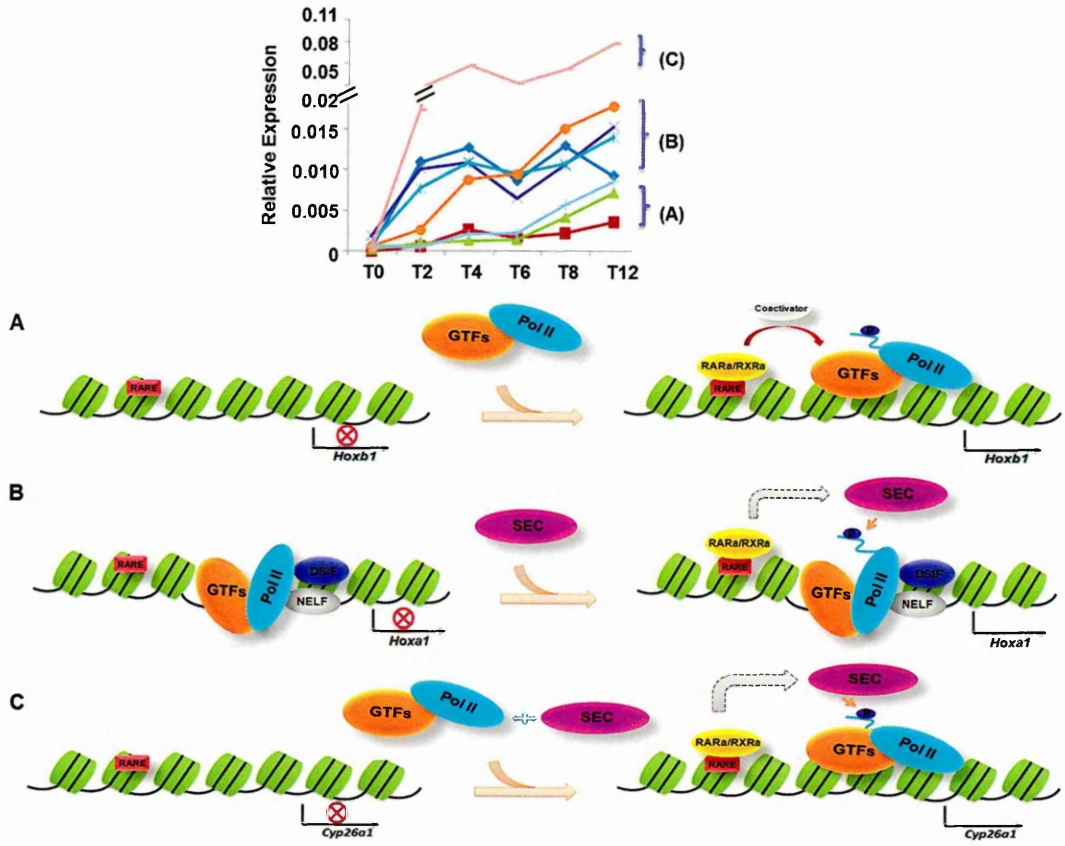


Figure 4.15. Diverse mechanisms for rapid activation of genes during development.
Top panel shows that rapidly activated genes can be further subdivided into distinct categories, A-C. (A) The *Hoxb1* gene newly recruits Pol II and general transcription factors (GTFs) in a classical gene activation mechanism, where RAR/RXR binds in the presence of RA, and with the help of coactivators, recruits GTFs and Pol II. (B) Paused Pol II, with DSIF/NELF, is present at the TSS of developmentally regulated genes, such as *Hoxa1*. In the presence of RA, RAR/RXR recruits SEC to stimulate transcription elongation through phosphorylation of the DSIF/NELF and the Pol II CTD. (C) *Cyp26a1*, a developmentally regulated gene that lacks paused Pol II, is induced by RA in a SEC-independent manner. All of the same factors are present after RA treatment as seen at *Hoxa1*, but *Cyp26a1* is induced to higher levels, suggesting that paused Pol II may serve to help regulate activation to equivalent levels.

Many developmentally regulated genes are marked by the presence of bivalent histone marks, the methylation of H3K4 and H3K27, DSIF/NELF and paused Pol II at the

TSS (Bernstein et al., 2006; Rahl et al., 2010; Stock et al., 2007). Since P-TEFb complexes, such as the SEC, are proposed to release paused Pol II via phosphorylation of the CTD and other general factors within the transcription complex, I asked whether SEC is recruited to these genes after induction of differentiation. I first focused on the *Hox* loci, because misregulation of *Hox* transcription is strongly implicated in leukemogenesis by MLL chimeras. Although a large number of developmentally regulated genes contain bivalent marks and paused Pol II at their promoters, I found that only a subset of *Hox* genes followed this pattern. Importantly, after induction of differentiation, *Hoxa1* was induced more rapidly than its paralog *Hoxb1* (Figure 4.15A-B). The SEC was specifically recruited to *Hoxa1*, and not *Hoxb1*, suggesting that SEC releases paused Pol II for rapid induction of transcription during development. This mechanism helps to explain the more rapid induction and regulatory roles of *Hoxa1* compared with *Hoxb1* in early neural development (Alexander et al., 2009). Additional examples of rapidly induced genes bearing paused Pol II at their promoter-proximal region that also recruited SEC were also identified in this study; and many of these were among the most rapidly induced. These findings were shown to be more general by studying the recruitment of SEC to the immediate early genes in HCT-116 cells after serum induction.

The *HSP70*, *FOS*, *JUN*, and *EGR* families of genes are well-studied, rapidly induced, and contain paused Pol II in the unstimulated state, leading to the paradigm that rapid induction is the primary function of paused Pol II (Donner et al., 2010; Nechaev and Adelman, 2008). However, paused Pol II is not present on *Cyp26a1* before its rapid induction to high levels of transcription, which suggests that paused Pol II is not a prerequisite for rapid induction, but rather facilitates coordinated and controlled induction. Studies in *Drosophila* have shown that developmentally regulated genes that have paused Pol II are activated in a synchronous manner, while developmentally regulated genes that

lack paused Pol II have a more stochastic pattern of induction during development (Boettiger and Levine, 2009; Levine, 2011). Having preloaded Pol II and GTFs reduces the number of steps for productive transcription, and thus, could result in a more equivalent and uniform way to induce gene expression. Genes such as *Cyp26a1*, while being required for proper development and being induced rapidly to high levels, may not need to be as precisely regulated at the earliest time points of induction.

Overall, our studies demonstrate that SEC is involved in many of the rapid and dynamic inductions of gene expression responses to developmental and environmental cues. P-TEFb was identified over 15 years ago (Marshall and Price, 1995) and was soon shown to be required for HIV transactivation (Mancebo et al., 1997; Wei et al., 1998; Zhu et al., 1997). Although the majority of P-TEFb is in the inactive Hexim1 complex, P-TEFb has also been shown to associate with a variety of factors that could help recruit it to chromatin in an active form (Bres et al., 2008; He and Zhou, 2011). In this manuscript, we have shown that the recently discovered SEC version of P-TEFb is a major regulator of rapidly induced genes in development. However, our genome-wide analyses indicate that not all rapidly activated genes require the SEC components investigated in this study. How the different P-TEFb complexes are recruited to regulate distinct sets of genes will be an important area of future investigations.

Chapter 5. The RNA Pol II Elongation Factor Ell3 Marks Enhancers in ES cells and Primes Future Gene Activation

5.1. Introduction

A full molecular understanding of how transcriptional networks are regulated in a pluripotent stem cell, resulting in coordinated differentiation into a complex organism, remains as one of the greatest challenges in biology. Enhancers play pivotal roles in modulating gene expression in a spatially and temporally specific pattern during development and are renowned for their ability to communicate with their associated genes from great distances. Physical interactions between an enhancer and a promoter have been proposed to explain how an enhancer influences gene activation (Bulger and Groudine, 2011). Several factors including cohesin and the CCCTC-binding factor (CTCF) have been shown to be involved in this process. CTCF can serve either as a barrier to protect a gene from position effect variegation or as a blocker to prevent long-range enhancer-promoter interactions (Engel et al., 2004; Noonan and McCallion, 2010; Wendt and Peters, 2009).

Cohesins have been shown to form a ring-like structure to hold segments of DNA together and were originally discovered for their role in chromosome segregation during mitosis and meiosis (Dorsett, 2011; Haering et al., 2002; Nasmyth and Haering, 2009; Skibbens, 2009). Cohesin-mediated long-range DNA interactions in transcriptional regulation were first proposed in *Drosophila* as a result of screening for factors affecting the *cut* gene's interaction with its *ca.* 85 Kb distal enhancer (Rollins et al., 1999). Since the discovery of its role in this process, cohesins have been shown to be required for long-range DNA interactions at the *IFNG*, *H19/Igf2*, apolipoprotein, and beta-globin loci from *Drosophila* to human (Dorsett, 2011; Hadjur

et al., 2009; Hou et al., 2010; Mishiro et al., 2009; Nativio et al., 2009). Indeed, most of what we know about cohesin and gene expression involves the interplay of cohesin and CTCF. Recently, a second class of cohesin sites, without CTCF, was described (Kagey et al., 2010). The presence of Mediator together with cohesin at enhancers was proposed to help bridge interactions between enhancers-bound transcription factors and RNA Pol II at the core promoter of active genes. These studies demonstrated the utility of defining specific classes of enhancers based on the cohort of bound factors.

In order to better predict and define the signatures of cis-regulatory elements and modifications functioning as enhancers, genome-wide sequencing analysis of genomic DNA and the analysis of chromatin occupancy and histone modifications have been used. The analysis of genomic DNA has focused on the identification of clusters of transcription factor motifs (Markstein and Levine, 2002) and resulted in the identification of highly occupied transcription (HOT) DNAs functioning as enhancers (Gerstein et al., 2010; Kvon et al., 2012; Moorman et al., 2006). Reduced nucleosome occupancy has also been used as a signature for enhancer identification (Khoueiry et al., 2010). Additionally, genome-wide chromatin modification studies have uncovered possible signatures for identifying enhancers. Over 100,000 putative enhancers can be identified in the human genome by combining the histone modifications and transcriptional coactivator, p300 (Creyghton et al., 2010; Heintzman et al., 2009). For example, the presence of p300, H3K4me1, and H3K27ac is proposed to mark active enhancers, whereas p300 and H3K4me1 alone, or with H3K27me3, marks poised or inactive enhancers (Creyghton et al., 2010; Heintzman et al., 2009; Rada-Iglesias et al., 2011).

Fundamental transcriptional studies over the past decade have pointed to the elongation stage of transcription as a major regulatory step in controlling gene expression (Levine, 2011; Sims et al., 2004; Smith et al., 2011a; Smith and Shilatifard, 2010). In embryonic stem (ES) cells, many developmentally regulated genes have paused Pol II at their promoters (Core et al., 2008; Guenther et al., 2007; Rahl et al., 2010). Many of these genes carry a bivalent chromatin mark consisting of both H3K4 and H3K27 trimethylation status (Bernstein et al., 2006; Mikkelsen et al., 2007). Recent studies classifying active and poised enhancers have shown that in ES cells, genes neighboring H3K27me3-marked enhancers are enriched for this bivalent chromatin mark and have lower expression than genes associated with active enhancers (Rada-Iglesias et al., 2011). Although the available genome-wide data has been successfully used to categorize enhancers and promoters into a limited number of predictive states, how these different classes of enhancers are used to regulate developmental gene expression is largely unknown.

Our studies have suggested that individual ELL family members can have distinct cellular roles. For example, in mouse ES cells, Ell2 has a prominent role as a component of SEC in the rapid yet synchronous activation of genes in response to retinoic acid signaling (Lin et al., 2011). In contrast, Ell1 is broadly required as a component of the Little Elongation Complex (LEC) in regulating snRNA gene expression in ES cells (Smith et al., 2011a). In order to define the genomic target specificity of Ell3, I analyzed its genome-wide occupancy in mouse ES cells. In this section, I show a unique role for Ell3 at enhancers for the regulation of gene activation during stem cell specification. I find that Ell3 occupies enhancers that are in a poised, active, or inactive state. Ell3 has an essential role in setting up paused Pol II at

developmentally regulated genes in a cohesin-dependent manner, potentially priming these genes for later activation by Ell2 within SEC during development.

5.2. Results

5.2.1. Ell3 occupies enhancer regions in mouse embryonic stem cells.

In order to further understand the functional diversity of the Ell family of proteins, I mapped the genome-wide distribution of Ell3 by ChIP-sequencing in mouse embryonic stem cells. In contrast to the enrichment of Ell1 at snRNA genes and Ell2 at other highly transcribed genes, Ell3 is preferentially found at intergenic regions (Figures 5.1A). Examination of well-characterized enhancers, such as the *Sox2* and *left-right determination factor 1* (*Lefty1*) enhancers, shows co-occupancy of Ell3 with known enhancer-binding factors such as p300 and octamer-binding protein 4 (Oct4) (Figure 5.1B).

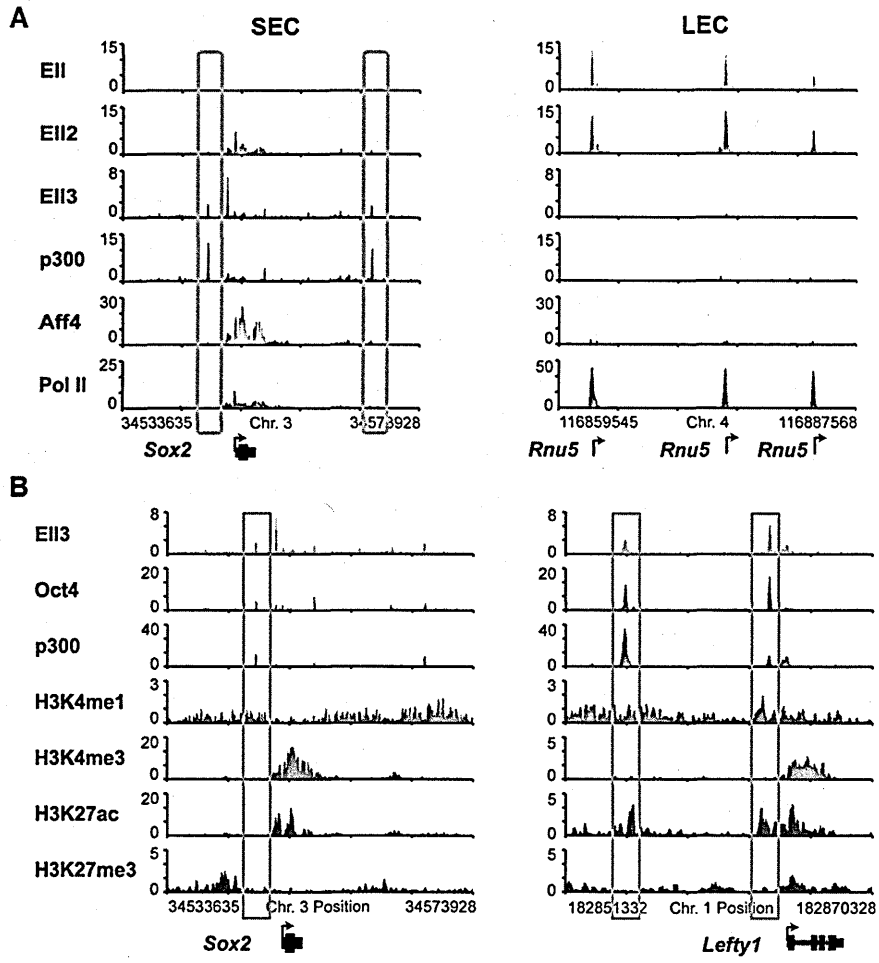


Figure 5.1. Ell3 co-localizes with p300 at enhancer regions.
(A) Distinct occupancy profiles of the three Ell family proteins. Ell3 co-localizes with p300 at promoter and distal regions of the actively transcribed *Sox2* gene (green boxes)(Creyghton et al., 2010). Ell2 colocalizes with its SEC partner AFF4 and RNA Pol II at the promoter and gene body of *Sox2*. Ell1 has a prominent presence, along with RNA Pol II, at three copies of the *Rnu5* snRNA genes. (B) Genome browser track examples for the occupancy profiles for Ell3, transcription factor Oct4, histone modifications (H3K4me1, H3K27ac, H3K4me3, and H3K27me3), and transcriptional coactivator p300 (Creyghton et al., 2010; Marson et al., 2008; Mikkelsen et al., 2007).

Genome-wide analysis identified 5,253 high-confidence regions bound by Ell3 with a False Discovery Rate (FDR) < 0.05 (Figure 5.2A). About 90% of these sites are enriched for enhancer markers such as p300 and H3K4me1 and only ~10% overlap with Transcription Start Site (TSS) regions occupied by H3K4me3 (Figures 5.2B and 5.2C), suggesting a global association of Ell3 with enhancers. Consistent

with this observation, an analysis of peak distributions shows that the majority of Ell3 peaks are located long distances (more than 10 Kb) from the core promoter regions (Figures 5.2D and 5.2E). Gene Ontology analysis of the nearest genes to the Ell3 peaks demonstrated that many of these genes are involved in developmental processes, including stem cell development, primary neural tube formation, embryonic pattern specification, and regulation of myeloid leukocyte differentiation (Figures 5.2F) (Huang da et al., 2009a, b; McLean et al., 2010).

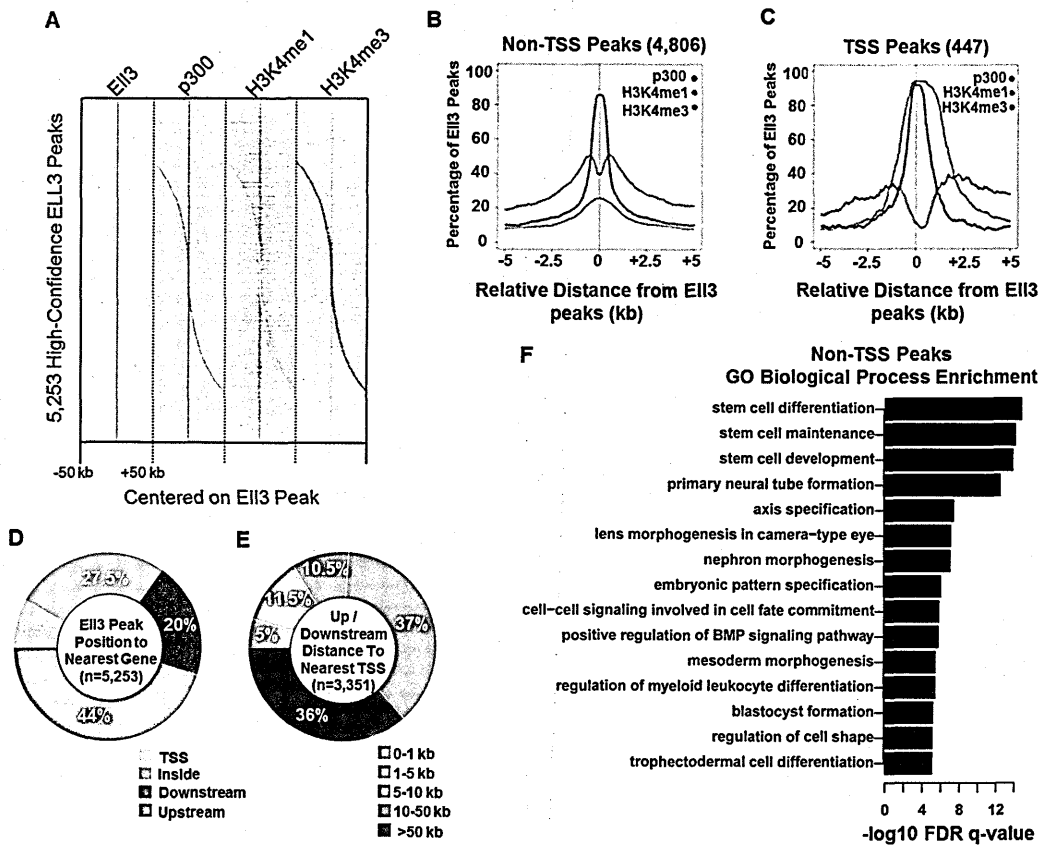


Figure 5.2. Ell3 predominantly associates with enhancer regions in mESC.

(A) Binding profiles for Ell3, p300, H3K4me1, and H3K4me3 are shown for regions 50 Kb upstream and downstream of all 5,253 high-confidence Ell3 peaks. Color indicates enrichment at FDR < 0.05. The majority of the Ell3-occupied regions are also enriched for the enhancer signature of p300 with H3K4me1, but not the Transcription Start Site (TSS)-associated H3K4me3 (Creyghton et al., 2010). (B-C) Profiles of p300, H3K4me1, and H3K4me3 centered on Ell3 peaks, shown 5 Kb around the Ell3 peak summit. Approximately 91.5% (4806) of all Ell3 peaks are found upstream or downstream of a TSS, and these are co-enriched for p300 and H3K4me1. In contrast, only 8.5% (447) of all Ell3 peaks are found at a TSS that is enriched for p300 and H3K4me3. (D) Pie chart showing that the percentages of Ell3 peaks that are upstream, downstream, within a gene or at a TSS. (E) Upstream and Downstream peaks were further categorized by their distance from the TSS. 73% of upstream/downstream Ell3 peaks are more than 10 Kb away from the nearest TSS. (F) Functional annotation of Ell3-bound non-TSS peaks, as reported by GREAT (McLean et al., 2010), indicates enrichment for developmental processes. The logarithmic x-axis values correspond to binomial FDR corrected –log10 q-values. The genome-wide data was done by the collaboration with Alexander Garrus.

5.2.2. Ell3 correlates with active, poised, and inactive enhancers.

Recent studies have indicated that genes proximal to H3K27ac-bound enhancers (active enhancer) have higher gene activity when compared with genes with H3K27me3-bound enhancers (poised or inactive enhancer) (Creyghton et al., 2010; Rada-Iglesias et al., 2011). My finding on the co-occurrences of Ell3 with H3K27ac on the active enhancers and with H3K27me3 on the poised or inactive enhancers suggests that Ell3 occupies both classes of enhancers (Figure 5.3). However, there are significant amounts of Ell3 and p300 bound enhancers that do not have either H3K27ac or H3K27me3 (Figure 5.3).

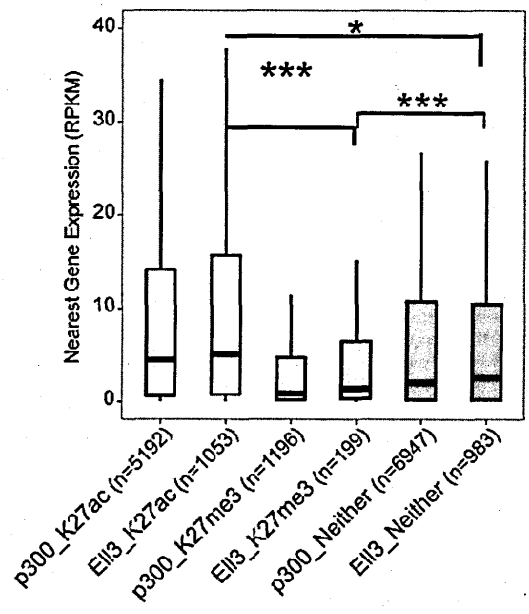


Figure 5.3. Ell3 is found at active, poised, and inactive enhancers.
Putative enhancers were selected for analysis based on the presence of p300, H3K4me1, and the absence of H3K4me3. Box plots show the expression levels of the genes proximal to enhancers with p300 and H3K27ac or H3K27me3 in the presence or absence of Ell3 (Creyghton et al., 2010). As previously shown for p300-bound enhancers, the presence of H3K27ac at Ell3-bound enhancers is associated with the higher expression of nearby genes, while the presence of H3K27me3 is associated with lower expression of nearby genes. There are 983 of the 2,235 Ell3-bound putative enhancers (44%) that have neither H3K27 acetylation nor H3K27 methylation. Putative Ell3 enhancers were defined as occurrence of p300 within 100 bp of an Ell3 peak and not within 100 bp of H3K4me3. Putative p300 enhancers were defined as occurrence of p300 not within 100 bp of H3K4me3. *, $p < 0.05$ and ***, $P < 0.0005$.

5.2.3. Ell3 is not required for stem cell self-renewal.

Since Ell3 occupies the active enhancer regions of the critical stem cell self-renewal genes Sox2, and also *Pou5f1* and *Nanog,* I sought to test whether Ell3 is required for stem cell self-renewal by regulating the expression of these genes. Lentivirus-mediated Ell3 shRNA was used to efficiently knock down the Ell3 levels in V6.5 ES cells (Figure 5.4A). Analysis on the mRNA levels of these self-renewal genes shows that the depletion of Ell3 does not have a major effect on their expression (Figures 5.4B). Furthermore, the formation of alkaline phosphatase positive ES colonies was not affected after Ell3 knockdown, suggesting that Ell3 is not essential for stem cell self-renewal (Figure 5.4C).

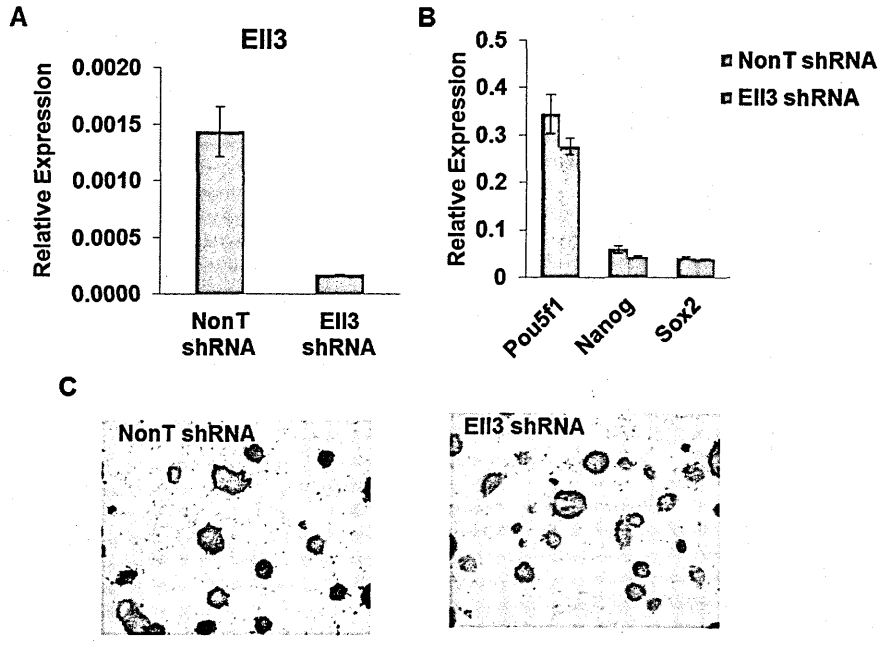


Figure 5.4. Ell3 is not required for the self-renewal of embryonic stem cells.
(A) qRT-PCR analysis of the Ell3 knockdown efficiency. (B) The expression levels of the stem cell self-renewal genes are largely unaffected after Ell3 knockdown. Expression levels were normalized to *Actin.* The error bar stands for the standard deviation of three independent measurements. (C) Alkaline Phosphatase (AP) staining of the control and Ell3 knockdown ES cells. ES cells were treated with non-targeting (NonT) and Ell3 shRNA for 72 hours before performing the AP staining.

5.2.4. Ell3 regulates the steady-state expression of a subset of neighboring genes.

To identify genes that are regulated by Ell3, I performed total RNA-sequencing analysis following shRNA-mediated Ell3 knockdown in mouse ES cells. There are 887 genes significantly down-regulated with a FDR < 0.05 and fold change > 1.5 in Ell3-depleted ES cells including *SRY-box containing gene 9* (*Sox9*) and *ST3 beta-galactoside alpha-2,3-sialyltransferase 1* (*St3gal1*).

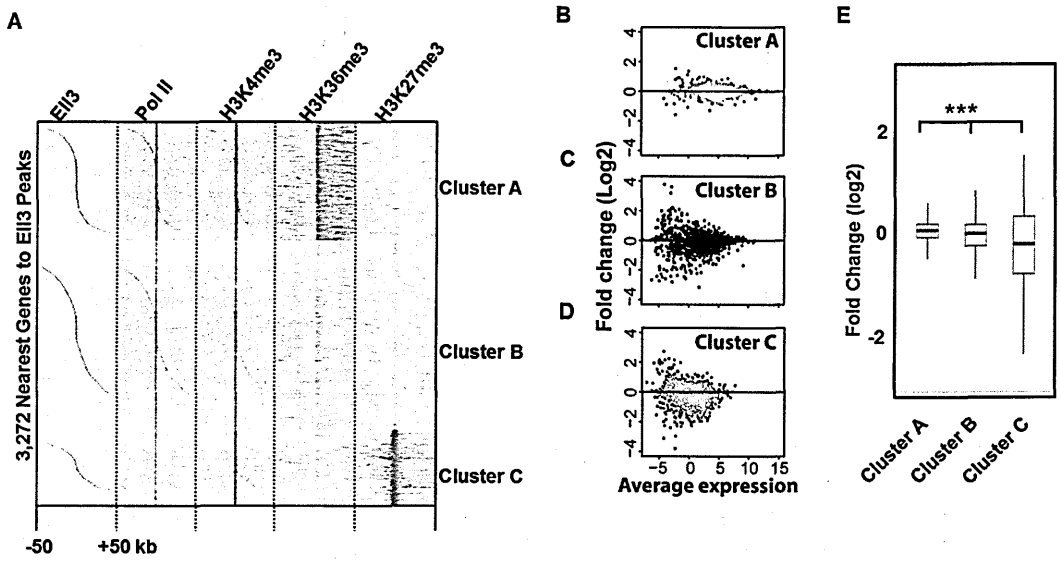


Figure 5.5. Ell3 preferentially regulates the expression of bivalent genes in ES cells.
(A) Cluster diagram of the 3,272 nearest genes to high-confidence Ell3 peaks. The non-redundant set of Ell3-associated genes was K-means clustered into three groups, **A**, **B**, and **C**, which are mainly distinguished by the profiles of H3K36me3 and H3K27me3 (Marson et al., 2008; Mikkelsen et al., 2007). The enriched ChIP-Seq signals for Ell3, Pol II, and the histone modifications are shown within 50 kb around the TSS of these genes. Each line represents a gene, and color indicates enrichment. Clusters are sorted based on the position and minimum distance of Ell3 to the nearest TSS. All gene orientations are 5' to 3'. (B-D) MA plots show the differential expression of Cluster **A-C** genes in Ell3-depleted ES cells vs. control cells. The y-axis (M) of each plot is the log2 fold change of gene expression levels of RNAi over wild-type; the x-axis (A) of each plot is the log2 average fragment per million reads per kb of exon as reported by Cufflinks. (E) Gene expression analyses of control (NonT) and Ell3-depleted ES cells. Genes in Cluster **C**, but not Cluster **A** and Cluster **B**, show a significant decrease in expression after Ell3 knockdown. Only genes with statistically sufficient coverage by RNA-seq are included in these plots (see Methods). The box indicates the middle quartiles of the distribution; the line indicates the median value, and the whiskers span 1.5 times the inter-quartile distance. *, $P < 0.05$; **, $P < 0.005$; and ***, $P < 0.0005$. The genome-wide data was done by the collaboration with Alexander Garrus.

In order to investigate whether Ell3 directly regulates the expression of specific classes of genes, we clustered the genes nearest to high-confidence Ell3 peaks based on the association of enrichment for Pol II and the histone modifications H3K4me3, H3K36me3, and H3K27me3 within 50 kb of the TSS of the nearest Ell3 genes. The 3,272 genes proximal to Ell3-occupied peaks were clustered into three major classes, **A-C** (Figure 5.5A). Cluster **A** (or "Active") genes, such as *Nanog* and

Sox2 (Figures 5.6A and 5.1B), show enrichment of Pol II and the active transcription marks H3K4me3 and H3K36me3, and have the highest transcription levels (Figures 5.5A-5.5E). Cluster **B** (or "Basal") genes, such as *arginine glutamic acid dipeptide (RE) repeats* (*Rere*) (Figure 5.6B), are characterized by low or no detectable levels of histone modifications H3K36me3 and H3K27me3, and low transcription (Figures 5.5A-5.5E). Cluster **C** genes, such as *St3gal1* (Figure 5.6C), which are marked by H3K27me3, have an even lower "Constrained" expression level as assayed by RNA-seq (Figures 5.5A-5.5E). For all clusters, we only included genes with statistically sufficient coverage by RNA-seq for the expression analysis in Figure 5E. Interestingly, fold expression levels of Cluster **C** genes are the most significantly down-regulated in Ell3-depleted cells (Figure 5.5E). Compared with the fold change of Cluster **A** genes, Cluster **B** genes do show slight reduction, which is statistically significant ($P < 0.0005$) (Figure 5.5E). Consistent with this observation, MA-plots also show the most significant reduction of the Cluster **C** genes after Ell3 knockdown and a slight reduction of Cluster **B** genes, but no significant effect on Cluster **A** genes (Figures 5.5B-5.5D). Therefore, enhancer-associated Ell3 mainly affects the expression of a subset of the "constrained" genes in mouse ES cells with a subtle effect on the basal expression of Cluster **B** genes.

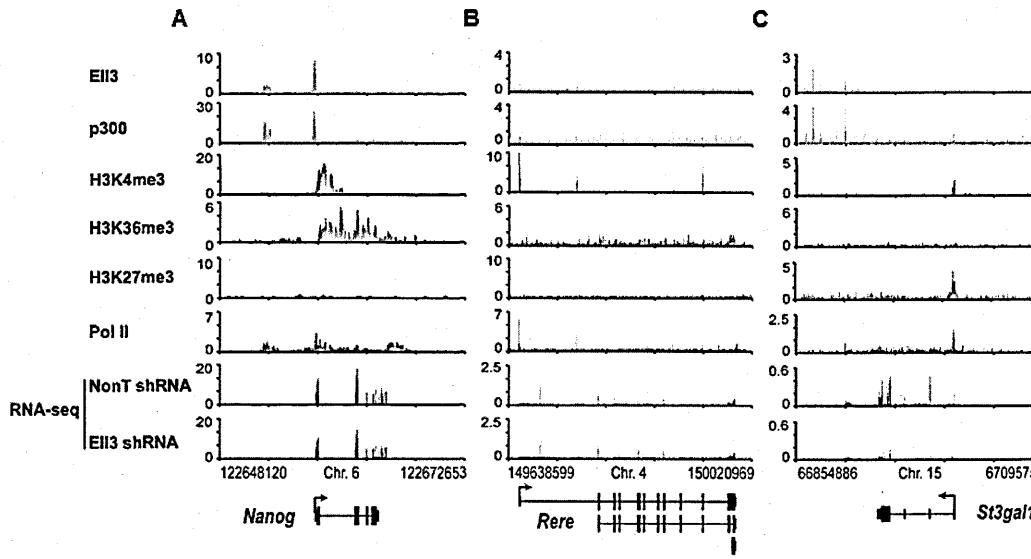


Figure 5.6. Examples of Ell3-associated genes in ES cells.

(A-C) Genome browser track examples of Groups **A-C** genes. Ell3 co-localizes with p300 at enhancer regions (blue box). RNA-seq analysis (purple box) shows reduced expression of the Group **C** gene, *St3gal1* (the RPKM values: 2.99 in Control and 0.62 in Ell3 knockdown, P=0), upon Ell3 knockdown.

5.2.5. Ell3 is required for the activation of genes during stem cell differentiation.

Many of the bivalent genes can be activated during differentiation (Bernstein et al., 2006). Since the Ell3-bound cluster **C** is enriched for bivalently marked genes, I asked whether their induction requires Ell3. Differentiated, day-5 embryoid bodies (EBs) were derived from control and shRNA-mediated Ell3 knockdown of ES cells for three, five, and ten day periods. As shown by quantitative RT-PCR analyses, the activation of *Sox9*, *iroquois homeobox 3* (*Irx3*), *St3gal1*, and *forkhead box P2* (*Foxp2*) were significantly reduced in the Ell3-depleted EBs, especially in day 5 and day 10 EBs (Figures 5.7A-5.7D). I also notice that the size of embryoid bodies formed from the Ell3-depleted ES cells is smaller than embryoid bodies formed from the control. Thus, apart from regulating the constrained expression of its proximal Cluster **C** genes, Ell3 is also required for their further activation during development.

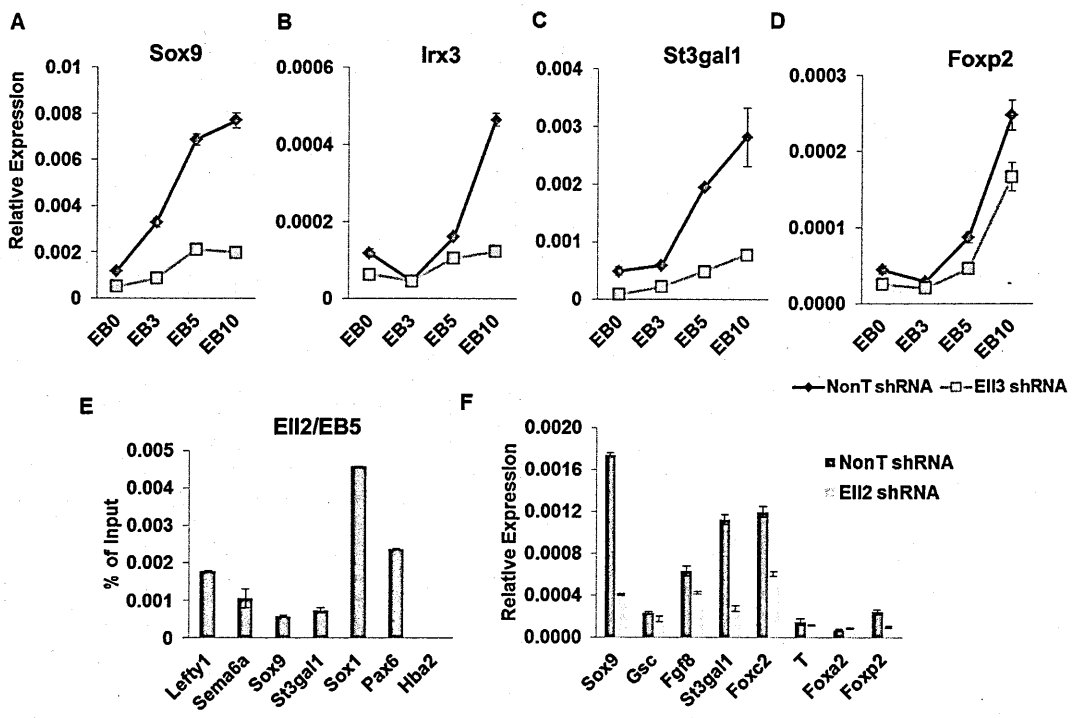


Figure 5.7. Ell3 and Ell2 are both required for the activation of bivalent genes associated with Ell3-occupied enhancers.

(A-D) qRT-PCR analyses of the activation time-course of four bivalently marked genes in the control and Ell3 knockdown EBs. Control and Ell3 knockdown ES cells were induced to form EB for 0 (EB0), 3 (EB3), 5 (EB5), and 10 (EB10) days, as indicated. The expression levels were normalized to *Actin*. All four genes showed reduced induction kinetics in the absence of Ell3. (E) Ell2 is recruited to the promoters of Ell3-regulated genes in 5-day EBs, as shown by ChIP. The *Hba* gene serves as a non-transcribed control gene. (F) Ell2 is required for the activation of many of the genes regulated by Ell3 in EBs. The control and Ell2 knockdown ES cells were induced to form EBs in the petri dishes for 5 days before the qRT-PCR analyses. The expression levels were normalized to *Actin*. The error bar stands for the standard deviation of three independent measurements.

5.2.6. Ell2 is also required for the activation of some of the Ell3 responsive genes.

Previously, I have demonstrated that Ell2 within SEC plays important roles in the rapid induction of several developmentally regulated genes in ES cells (Lin et al., 2011). Many of these Ell2/SEC-responsive genes contain engaged Pol II at their promoter-proximal regions in the undifferentiated state. Therefore, I asked if Ell2/SEC is required for the activation of Ell3-regulated genes. Chromatin

immunoprecipitation in differentiated day-5 EB samples shows that Ell2 is indeed newly recruited to the promoter regions of the *Sox9* and *St3gal1* genes (Figure 5.7E). Furthermore, qRT-PCR analysis of day 5 Ell2-depleted EBs revealed that Ell2 is also required for the activation of many Ell3-responsive genes, including *Sox9*, *St3gal1,* and *Foxp2* during development (Figure 5.7F). Taken together, these results suggested that Ell3 might function upstream of Ell2/SEC in transcriptional programs during development.

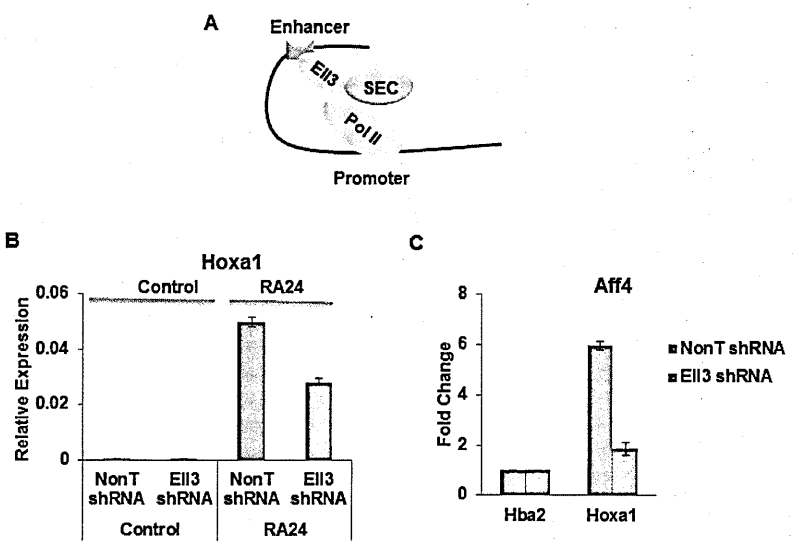


Figure 5.8. Ell3 binding at enhancers is required for future gene activation by SEC.

(A) Schematic model for Ell3 pre-binding at enhancers primes future gene activation by SEC. (B) Ell3 binding to enhancers is required for the activation of *Hox* genes by retinoic acid (RA). Control and Ell3 knockdown ES cells were untreated (Control) or treated with RA for 24 hours (RA24) before harvesting for the qRT-PCR analysis. (C) Ell3 is required for the recruitment of SEC (Aff4) to the *Hoxa1* gene after RA treatment. ChIP signal is normalized to the non-transcribed *Hba2* gene. Error bars indicate the standard deviation of three independent measurements.

5.2.7. Ell3 binding at enhancers is required for future gene activation by SEC.

To further investigate whether the pre-binding of Ell3 to enhancers is essential for the future gene activation through the recruitment of SEC, we first examined the

requirement of Ell3 for the activation of *Hoxa1* by retinoic acid (RA), a gene activated by SEC (Lin et al., 2011) (Figure 5.8A). Quantitative RT-PCR analysis indicates that the activation of *Hoxa1* by RA is reduced after Ell3 knockdown (Figure 5.8B). In addition, the depletion of Ell3 impairs the recruitment of Aff4, the central factor of SEC, to the *Hoxa1* promoter after 24 hours of RA treatment (Figure 5.8C). Our previous biochemical studies indicated that ELL3 can interact with AFF4 and P-TEFb to form a complex similar to ELL2-containing SEC in 293 cells (Lin et al., 2010). We have also found that Ell3 can interact with Aff4 and P-TEFb when overexpressed in ES cells. Therefore, we propose that Ell3's binding to enhancers is required for the full assembly of SEC on the promoter of genes, and thus, future gene activation by SEC at the Ell3 target genes.

5.2.8. Ell3 binding at enhancers regulates the Pol II occupancy at promoter-proximal regions of neighboring genes.

Ell3 was initially identified as a member of the ELL family of RNA Pol II elongation factors, which can increase the transcription elongation rate catalyzed by Pol II (Miller et al., 2000). ELL family members can directly interact with Pol II, and are proposed to facilitate the proper alignment of the 3' terminus of the nascent transcript with the Pol II active site (Elmendorf et al., 2001; Shilatifard et al., 2003). Since Ell3 binding to enhancers is required for the proper expression of a subset of Group **B** and **C** genes (Figure 2), we asked whether Ell3 is required for the proper occupancy of Pol II at its nearest genes by performing Pol II ChIP-seq after Ell3 knockdown. Loss of Ell3 leads to reduction of Pol II in many Group **C** genes, such as *St3gal1*, and Group **B** genes, such as *Rere*, with a lesser effect on Group **A** genes, like *Nanog* (Figures 5.9A-C).

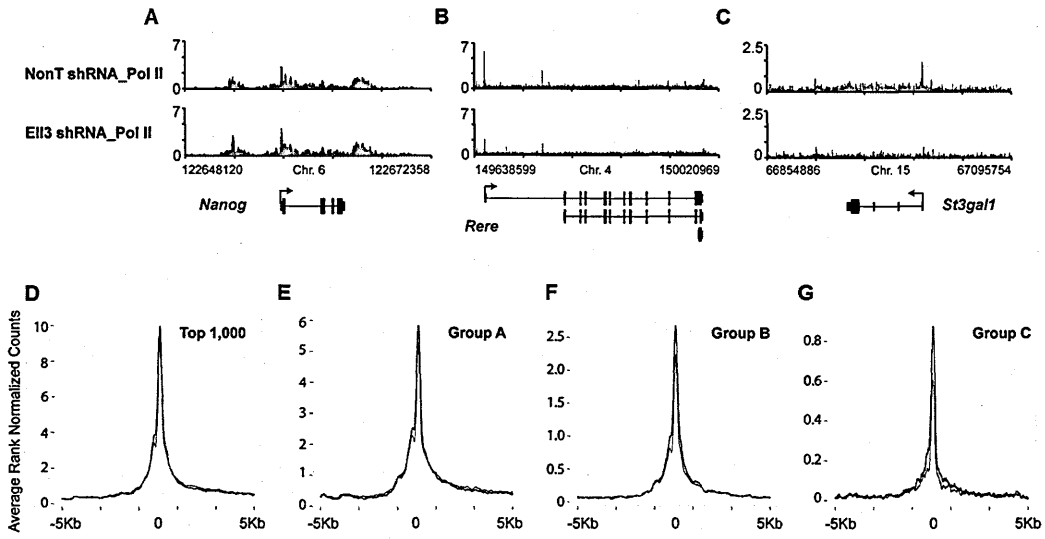


Figure 5.9. Ell3 regulates Pol II occupancy at promoter-proximal regions of neighboring genes. (A-C) Genome browser profiles of Pol II occupancy in control and Ell3-depleted cells. Pol II levels are reduced at the *Rere* and *St3gal1* genes, but not the *Nanog* gene. (D-G) Average Pol II occupancy plots for the top 1,000 highly expressed genes and Ell3 nearest genes from the Figure 2A group analysis. Rank normalized average Pol II levels within 5 kb of the TSS are shown in control (black line) and Ell3 knockdown (red line) ES cells. Pol II is reduced at the TSS region of Ell3-associated genes, with strong effects on Group **C** genes.

In order to further investigate whether Ell3 differentially regulates the Pol II occupancy in Groups A-C genes genome-wide, we directly compared the occupancy levels of Pol II at promoter-proximal regions of genes nearest to Ell3-bound peaks in control and Ell3-depleted ES cells. Compared with Group **A** and **B**, Group **C** genes show the largest reductions in Pol II occupancy (Figures 5.9D-G). We note that group C genes are expressed at a very low level, and it therefore may be easier to observe a larger fold change in Pol II occupancy and expression after Ell3 RNAi than at highly expressed genes. However, our data suggest that, during the process of gene activation, genes may achieve a state at which they no longer require Ell3 at their enhancers for the maintenance of expression.

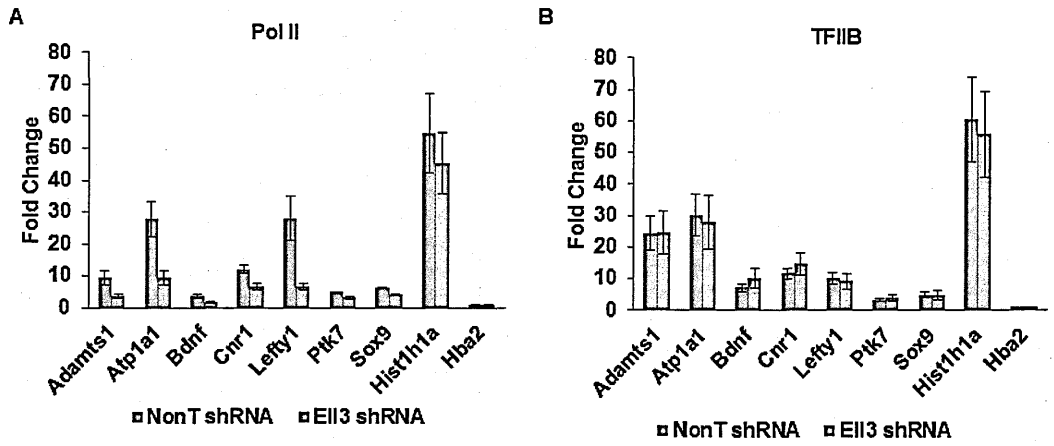


Figure 5.10. The recruitment of basal transcription factor TFIIB is not affected by Ell3 knockdown.

A) Knockdown of Ell3 reduces the promoter-proximal Pol II occupancy at many bivalent genes, but not the actively transcribed histone H1 gene, *Hist1h1d.* The *Hba2* gene serves as a non-transcribed control gene. (B) Ell3 depletion does not affect the recruitment of the basal factor TFIIB. The *Hba2* gene serves as a non-transcribed control gene. The error bars represent the standard deviation of three independent measurements.

To rule out the possibility that Ell3 might affect the assembly of the basal transcriptional machinery at an early stage, the levels of the basal factor TFIIB loading was also examined in Ell3 knockdown cells (Figures 5.10A and 5.10B). I observe that the basal transcription factor TFIIB is properly recruited to the promoter regions of bivalent genes upon Ell3 depletion. Thus, Ell3 is essential for the establishment of promoter-proximal pausing of Pol II at many genes in ES cells.

5.2.9. Ell3-dependent promoter-proximal pausing requires the cohesion complex.

One of the well-accepted models for the regulation of neighboring gene activities by enhancers is promoter-enhancer looping (Bulger and Groudine, 2011; Li et al., 2012; Lieberman-Aiden et al., 2009; Montavon et al., 2011). It has recently been shown that the cohesin and Mediator complexes occupy both the enhancer and promoter regions,

promoting loop formation between the enhancer - promoter pairs at active genes (Kagey et al., 2010). Depletion of the Mediator or cohesin complexes reduces the interactions between the enhancer and core promoter of the *Nanog* gene in ES cells (Kagey et al., 2010). Interestingly, the cohesin and Mediator complexes are found on many other Ell3-bound enhancers, including the *Lefty1* and *St3gal1*, as well as the hypersensitivity site 2 (HS2) enhancer of the beta globin locus, which is known to be regulated by cohesion in erythroid cells (Hou et al., 2010) (Figure 5.11). Beta globin genes are completely silent genes in ES cells, suggesting a possible role of the cohesin/Mediator complexes in enhancer-promoter communication at inactive or poised genes.

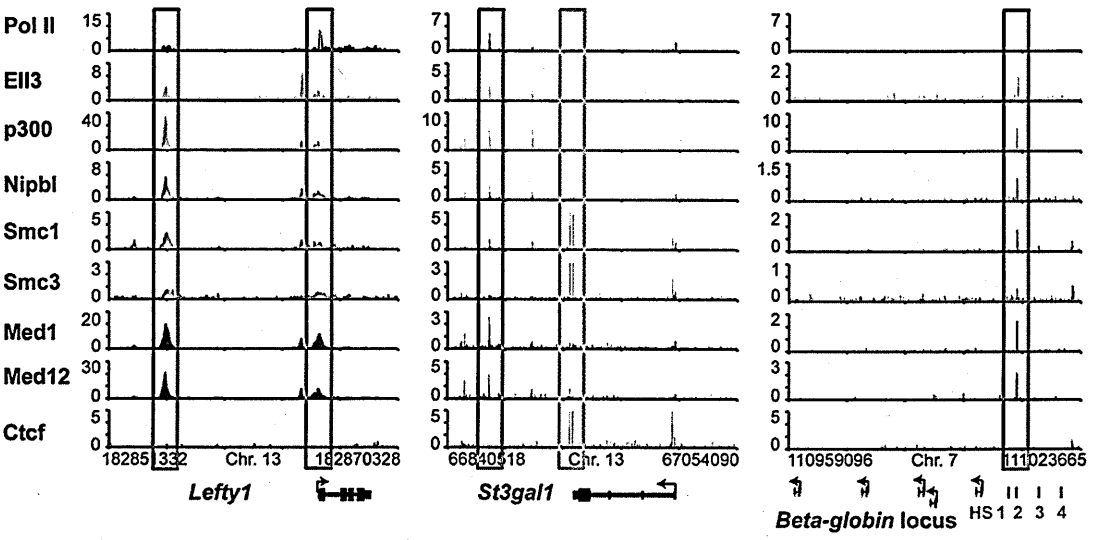


Figure 5.11. Ell3 co-occupies with cohesin and Mediator at enhancers.

Genome browser profiles for Ell3, p300, cohesin (Nipbl, Smc1a, and Smc3), Mediator components (Med1 and Med12), and Ctcf (Kagey et al., 2010). Ell3 is found to colocalize with cohesin at sites that are enriched for Mediator and have low Ctcf occupancy (blue boxes). Ell3 is not enriched at cohesin sites that have high Ctcf and low Mediator occupancy (green box). Ell3 colocalizes with cohesion and Mediator at the HS2 element of the inactive globin locus.

To explore whether the cohesin complex is also required for the proximal Pol II pausing at these genes, Pol II occupancy was first examined in cohesin subunit-depleted cells. The structural maintenance of chromosomes 1A (Smc1a) and structural maintenance

of chromosomes 3 (Smc3) proteins are significantly reduced by Smc1a and Smc3 shRNAs, respectively (Figure 5.12A). Smc3 knockdown also leads to a reduction of the protein levels of Smc1a, but not the Smc1a mRNA levels, suggesting that Smc3 affects Smc1a protein stability (Figures 5.12A). The depletion of the cohesin complex does not have much of an effect on global Pol II levels, whether looking at the total, phosphorylated, or unphosphorylated forms (Figure 5.12A). However, the Pol II occupancy at the promoters of the genes nearest to Ell3-bound peaks was greatly reduced after cohesin knockdown as shown by ChIP-qPCR (Figure 5.12B), indicating an important role of the cohesin complex in promoter-proximal pausing of Pol II at Ell3 target genes.

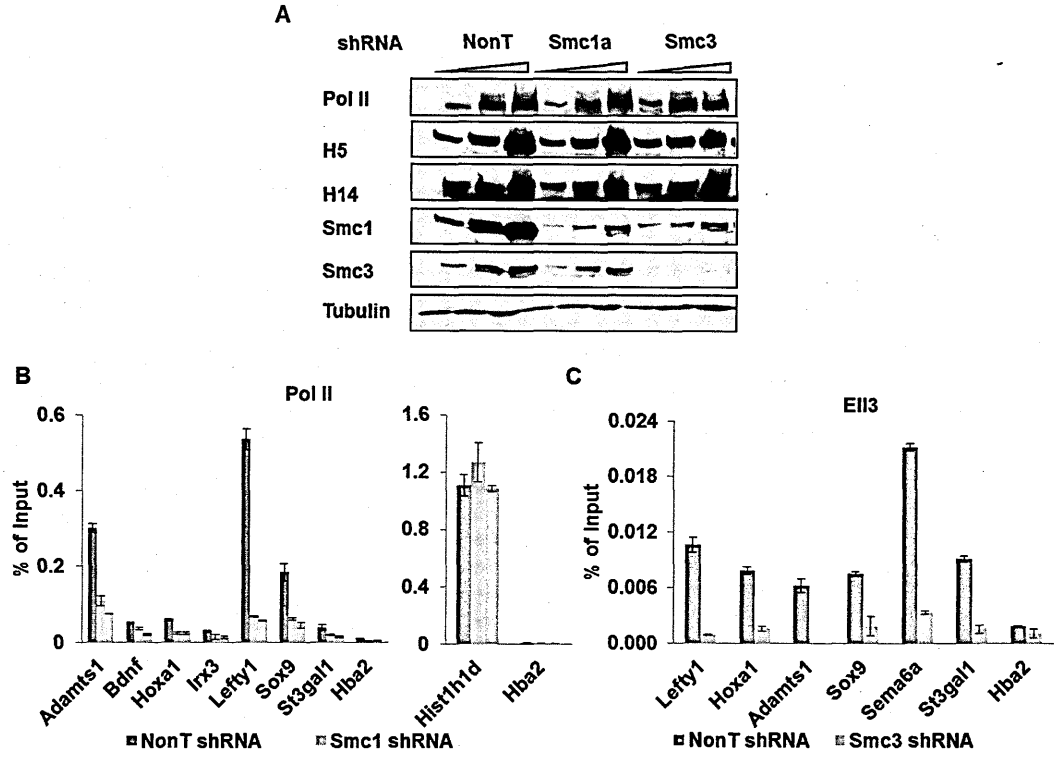


Figure 5.12. Ell3-mediated promoter-proximal pausing of RNA Pol II requires the cohesin complex. (A) Knockdown of cohesin components Smc1a or Smc3 does not affect cellular Pol II levels. Smc1a knockdown significantly reduces the Smc1a protein level, but not Smc3. However, the depletion of Smc3 also reduces the protein level of Smc1a. The unphosphorylated (8wg16 antibody), Ser5 phosphorylated (H14), and Ser2 phosphorylated (H5) forms of RNA Pol II levels remain unchanged upon the knockdown of cohesin components. Tubulin serves as a loading control. (B) Knockdown of cohesin components reduces the promoter-proximal Pol II occupancy at many Ell3-responsive genes. Histone H1 (*Hist1h1d*) and alpha globin (*Hba2*) serve as highly expressed and non-expressed control genes. (C) Knockdown of Smc3 reduces Ell3 occupancy at the enhancer regions of Ell3-responsive genes. The *Hba2* gene serves as a non-transcribed control gene. The error bar stands for the standard deviation of three independent measurements.

To further assess whether the role of Ell3 in promoter-proximal Pol II occupancy is mediated through the cohesin complex, we tested Ell3 binding profiles in the cohesin-depleted cells. The results revealed that depletion of the cohesin complex (Smc3 knockdown) also greatly reduces Ell3 occupancy at enhancer regions (Figure 5.12C). This is not due to the reduced expression of the *Ell3* gene, as quantitative RT-PCR shows that Ell3 mRNA levels remain unchanged after cohesin

knockdown (Figure 5.13). Therefore, the cohesin complex is required for the localization of Ell3 on these enhancers and for Ell3's effect on RNA Pol II at promoter-proximal regions.

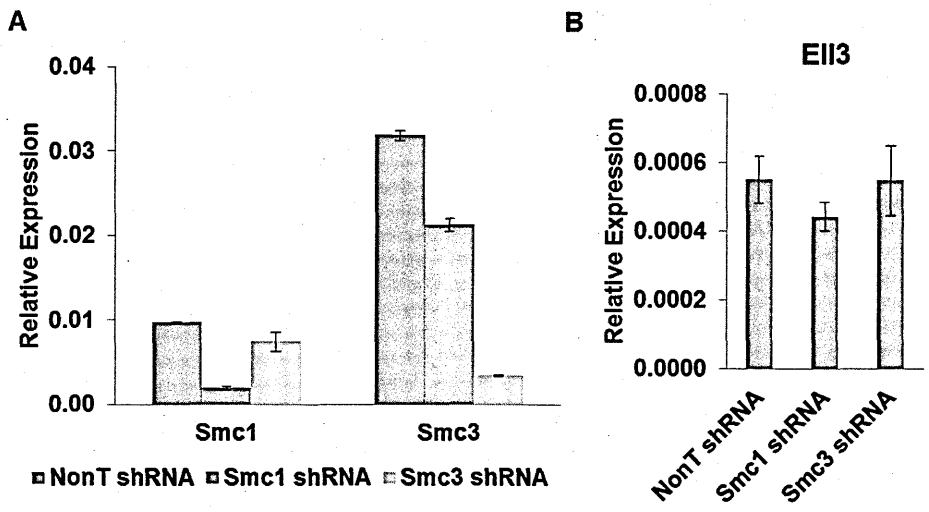


Figure 5.13. Cohesin is not required for the *Ell3* gene expression in mouse ES cells. (A) qRT-PCR analysis of the *Smc1* and *Smc3* knockdown efficiency. (B) The Ell3 mRNA level is not affected by the depletion of the Cohesin components. Expression levels were normalized to *Actin*. The error bar stands for the standard deviation of three independent measurements.

It has been suggested that the genome is spatially organized into many three-dimensional structures (Cremer and Cremer, 2010; Lieberman-Aiden et al., 2009). Recent studies have identified many local chromatin interaction domains, named "topological domains" (Dixon et al., 2012). Comparing genome-wide chromatin conformation data from ES cells with Pol II occupancy data shows that Pol II can be broadly down-regulated after Ell3 knockdown within structured chromosomal domains (Figures 5.14A-C). However, whether and how Ell3 can work through these structured domains to coordinately regulate several neighboring genes needs to be further investigated.

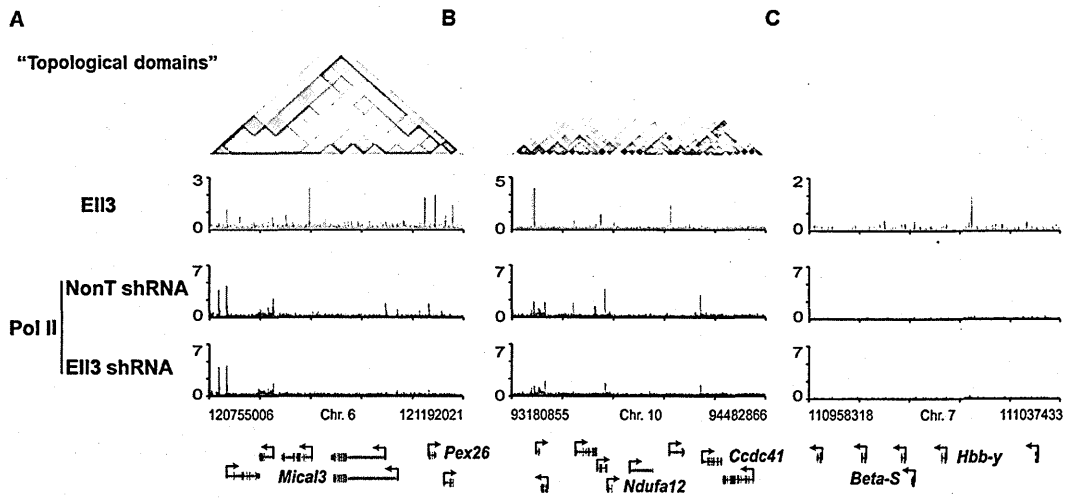


Figure 5.14. Pol II occupancy analysis after Ell3 knockdown in identified topological domains. (A-C) Genome browser tracks of Pol II within the defined three dimensional "topological domains" in control and Ell3-depleted cells. The deepness of the red color indicates the degree of intrachromosomal interactions by genome-wide chromosome conformation analysis (Dixon et al., 2012) (http://chromosome.sdsc.edu/mouse/hi-c/index.html, accessed on Jun 1st 2012). Ell3 occupies many putative and defined enhancer regions. Reductions of Pol II occupancy throughout the structured chromosomal domains were observed after Ell3 knockdown. The inactive beta globin locus, which has Ell3 at the HS2 enhancer, but which lies in a chromosomal region with low intrachromosomal interactions in ES cells, serves as a control.

To further explore if Ell3's binding at enhancers could stabilize enhancer-promoter interactions, we performed a Chromosome Conformation Capture (3C) assay by anchoring on an Ell3 binding site at the *Hoxa* locus (Figure 5.15). We observe a broad domain of interactions at the *Hoxa* locus in the ES cell state, consistent with the published Hi-C data reporting this region as encompassing a topological domain (Figures 5.15A-B) (Dixon et al., 2012). Interestingly, after 24 hours of RA treatment, these interactions become more local and specific, and this transition in interactions requires Ell3 (Figures 5.15C-D).

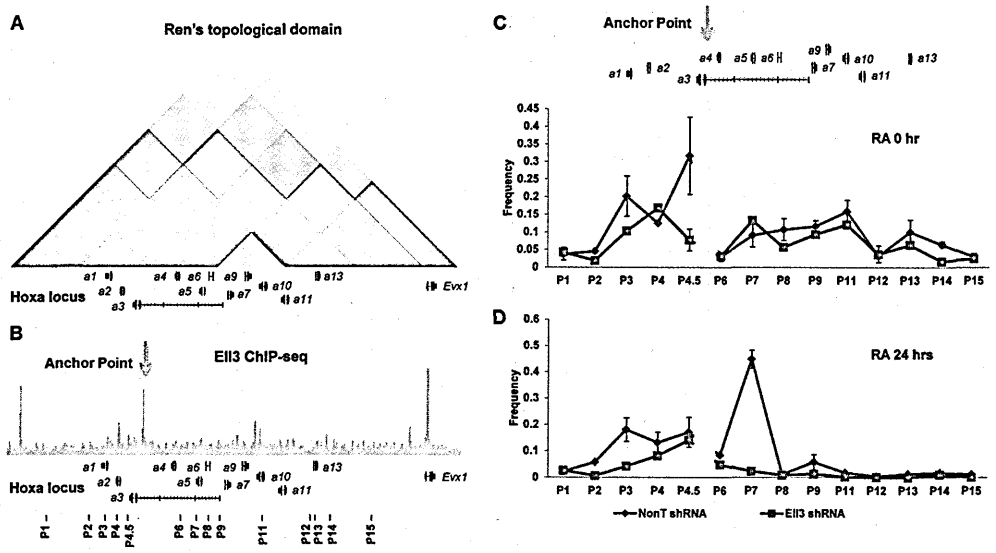


Figure 5.15. Ell3 is required for the specific enhancer-promoter interactions in *Hoxa* locus after induction.

(A) qRT-PCR analysis of the *Smc1* and *Smc3* knockdown efficiency. (B) The Ell3 mRNA level is not affected by the depletion of the Cohesin components. Expression levels were normalized to *Actb*. The error bar stands for the standard deviation of three independent measurements. (C) Genome browser track example for Hi-C results at the *Hoxa* locus (Dixon et al., 2012) (http://chromosome.sdsc.edu/mouse/hi-c/index.html, accessed on Sep 1st 2012). (D) Genome browser track example for the binding profile of Ell3 at the *Hoxa* locus. Primer regions used for the Chromosome Conformation Capture (3C) assay were labeled as P1-15. Green arrow marks the anchor point for the 3C assay. (E-F) Ell3 is required for the specific enhancer-promoter interactions after 24 hours of RA treatment. 3C results among different samples were normalized to a control locus Ercc3.

5.2.10. Ell3 is essential for stem cell pluripotency and differentiation.

Interestingly, in addition to the above-mentioned Group **A-C** genes, some inactive or "Dormant" lineage-specific genes are also associated with Ell3/p300-bound enhancers, but have no detectable Pol II, H3K4me3, and H3K36me3 in their transcription units. For example, Ell3 and p300 are present at the above-mentioned well-characterized HS2 enhancer element of the beta globin locus, which is silent in ES cells (Figure 5.11C). Therefore, we consider HS2-like enhancers to be in an "inactive/dormant state" or Group **D**.

To explore if Ell3 is required for differentiation of ES cells, we further tested the expression levels of lineage-specific markers in the Ell3 knockdown of EBs. Quantitative RT-PCR analysis shows that many endoderm markers (forkhead box A2 (*Foxa2*), GATA binding protein 4 (*gata4*), and GATA binding protein 4 (*Gata6*)), mesoderm markers (goosecoid homeobox (*Gsc*), brachyury (*T*), and fibroblast growth factor 8 (*Fgf8*)), and ectoderm specific genes (potassium voltage-gated channel subfamily C member 1 (*Kcnc1*), GLI-Kruppel family member (*Gli1*), and oligodendrocyte transcription factor 3 (*Olig3*)) are significantly down-regulated in the Ell3-depleted EB samples compared with the control EB samples (Figures 5.16A-5.16C).

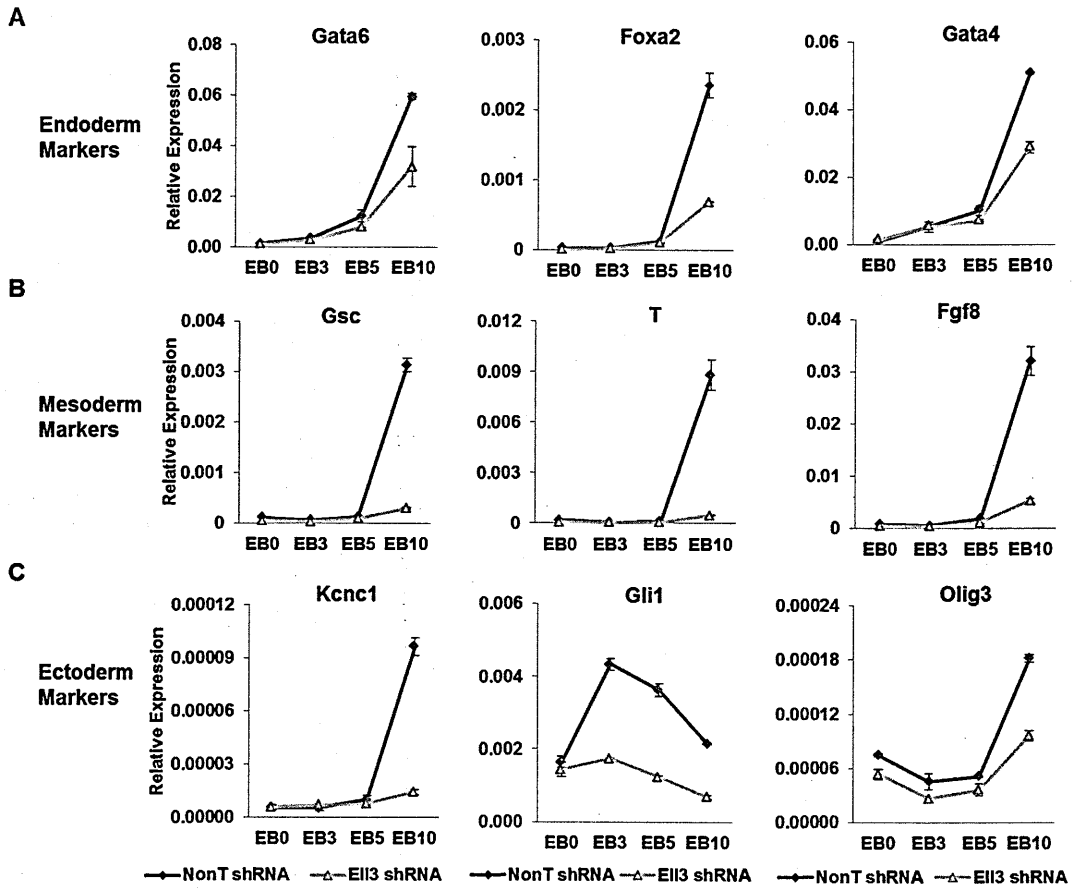


Figure 5.16. Ell3 is required for the stem cell specification.

(A-C) qRT-PCR analyses of the activation kinetics of lineage-specific genes in control and Ell3-knockdown EBs. Control and Ell3-knockdown ES cells were induced to form EB for the indicated time points. Expression levels were normalized to *Actin*. The error bar stands for the standard deviation of three independent measurements

Examination of Ell3 occupancy shows that Ell3 associates with many of the inactive or poised enhancers of these lineage-specific genes (Figure 5.17). Moreover, many of these genes do not contain detectable Pol II at their promoter-proximal regions (Figure 5.17). Further RNA-seq analysis on the differentiated day-5 EBs indicated that 2,862 genes were up-regulated with a FDR < 0.05 and fold change > 1.5. Of these, 510 of them contain Ell3 at their enhancers in the ES state suggesting the involvement of Ell3 in stem cell pluripotency.

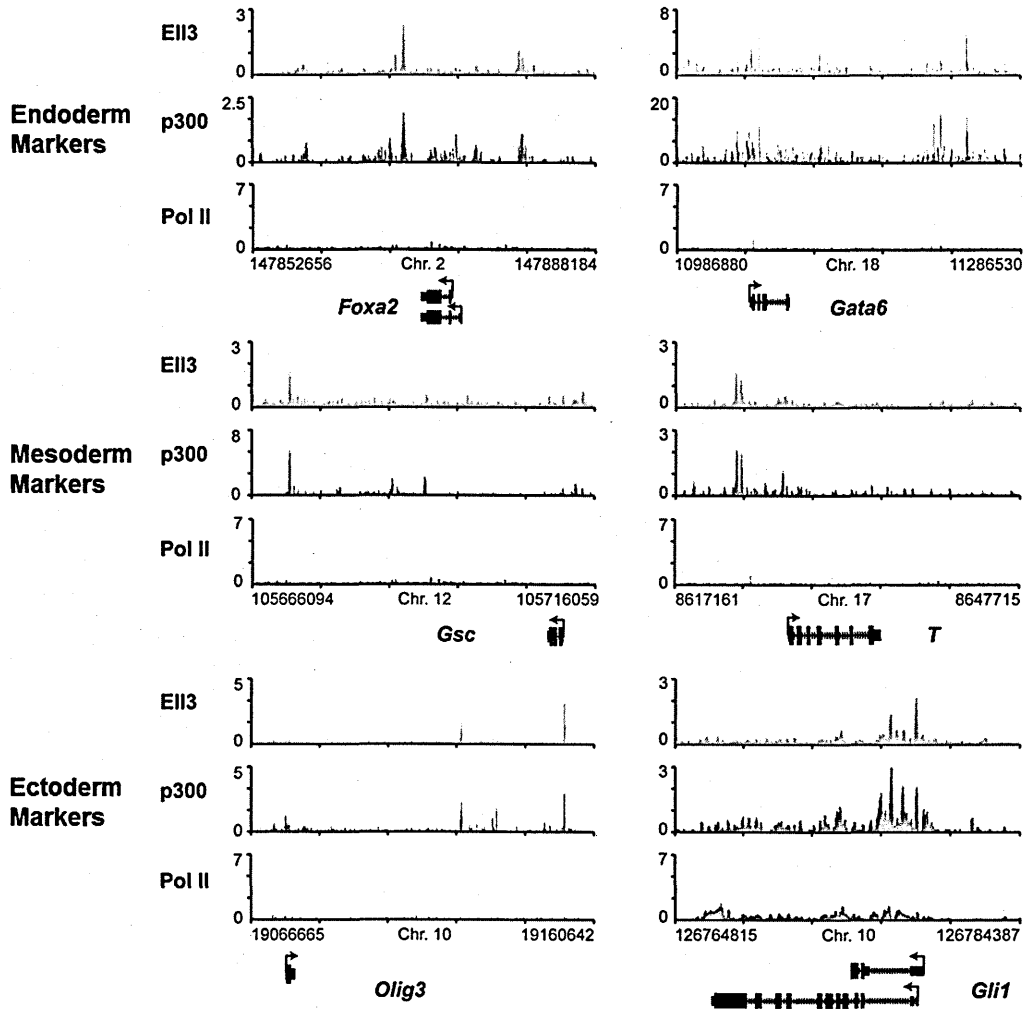


Figure 5.17. The binding profiles of Ell3, p300, and Pol II on lineage-specific genes.

Ell3 and p300 co-occupy enhancer regions of these lineage-specific genes. However, many of them, except *Gli1*, do not contain detectable Pol II at their promoter-proximal regions.

I next assessed the effect of Ell3 knockdown on neural induction by retinoic acid in ES-derived embryoid bodies. The β-tubulin III positive neural fibers are significantly reduced in the Ell3-depleted EBs compared with the control EBs (Figure 5.18). Taken together, these results suggest that although enhancer-associated Ell3 is not required for stem cell self-renewal, it is essential for stem cell pluripotency and differentiation.

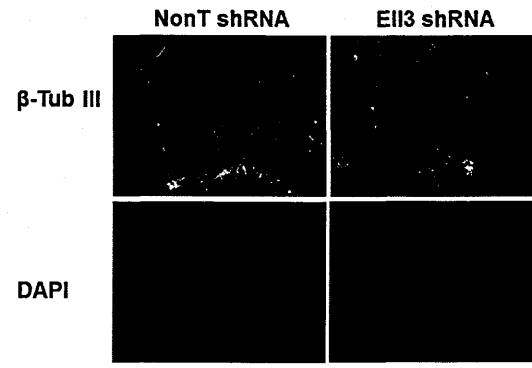


Figure 5.18. Ell3 is essential for the proper neural differentiation of mouse ES cells.
The 5-day differentiated EBs from control and Ell3-knockdown ES cells were further differentiated into neural cells by exposure to retinoic acid for 14 days. Neural-differentiation competence was visualized by immunostaining for the neuronal marker class III, β-tubulin (β-tub III, green), and the DNA marker, DAPI (blue).

5.2.11. Ell3 is present on the chromatin of germ cells.

The presence of Ell3 at enhancers in ES cells of many dormant lineage-specific genes (both Group **C** and Group **D** genes) raises the question of at what stage is Ell3 recruited to mark these enhancers. Interestingly, our previous northern blot analyses indicated that Ell3 is highly enriched in testes (Miller et al., 2000). I, therefore, performed immunofluorescence staining of mouse sperm and observe that Ell3 localizes to sperm nuclei (Figure 5.19A). Immunogold labeling of Ell3 in mouse sperm by Ell3 antibodies raised against either the N-terminus or the C-terminus of Ell3 further validates the nuclear localization of Ell3 in sperm (Figures 5.19B and 5.20). Interestingly, we also detect Pol II in sperm nuclei by immunogold labeling (Figure 5.19B). While Ell3 N- and C-terminal antibodies co-localize well (within 5-10nm) in sperm nuclei, antibodies to Ell3 and Pol II appear to occupy different regions on the sperm cell chromatin (Figures 519B).

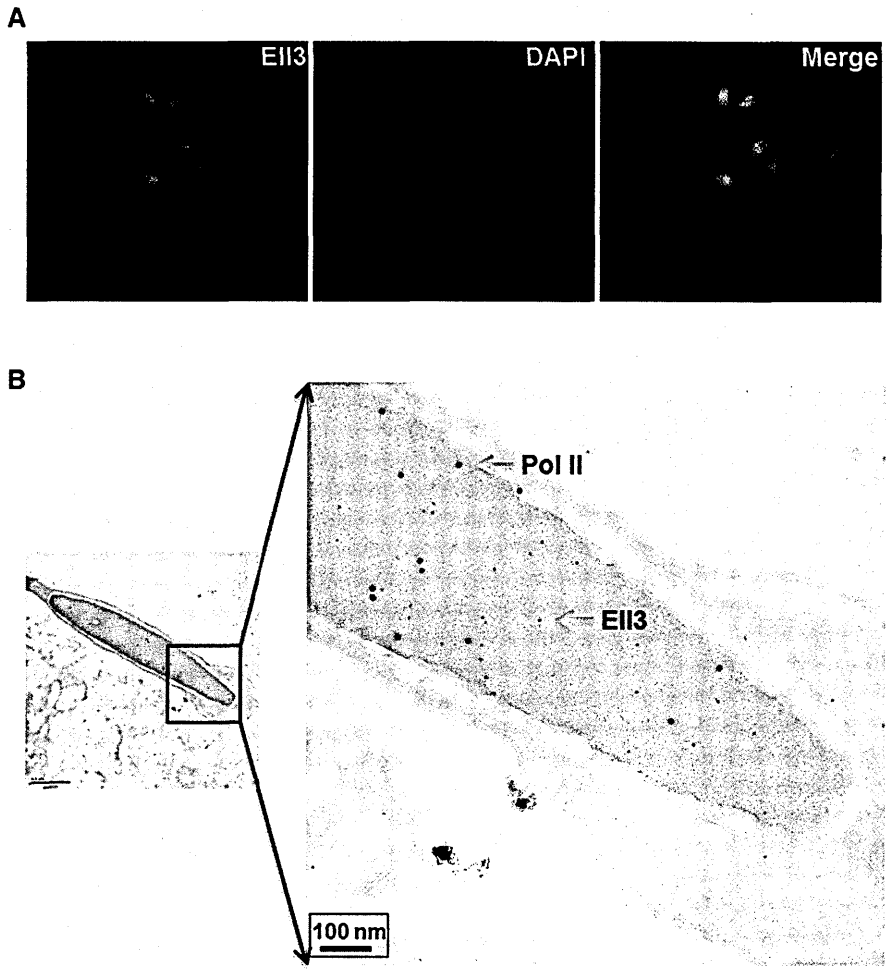


Figure 5.19. Ell3 and Pol II localization on the chromatin of germ cells.

(A) Immunofluorescence staining of Ell3 in mouse sperm. Mouse sperm were fixed and stained with antibodies raised against the C-terminus of mouse Ell3, and then were counterstained with DAPI. (B) Immunogold labeling of Ell3 and Pol II in mouse sperms. Mouse sperms were fixed, cryo-sectioned, and double stained with Ell3 N-terminus and Pol II antibodies. Both Ell3 (red arrow, 6 nm gold particles) and Pol II (blue arrow, 12 nm gold particles) localize in the nucleus of the sperm. Co-localization of Ell3 and Pol II was largely not observed compared to the co-localization of the N- and C-terminally raised Ell3 antibodies, which are frequently found within 5-10 nm of each other (Figure 5.20). The electron microscopy study was done by the collaboration with Fengli Guo from Histology core facility.

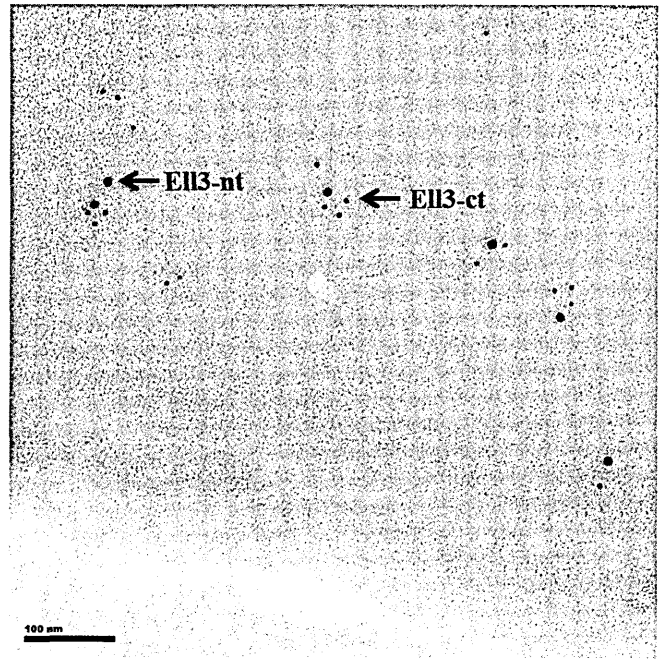


Figure 5.20. Immunogold labeling of Ell3 in mouse sperm

Mouse sperm were fixed, cryo-sectioned, and sequentially stained with Ell3 C-terminal (Ell3-ct) and N-terminal (Ell3-nt) antibodies. Ell3-ct (6 nm gold particles) and Ell3-nt (12 nm gold particles) co-localize in the nucleus of the sperm within 5-10 nm of each other. This figure was done by the collaboration with Fengli Guo from Histology core facility.

5.3. Discussion

Regulatory elements play a central role in establishing promoter-proximal engaged Pol II, as previously demonstrated for the *Drosophila Hsp70* gene and the murine *Ig kappa* gene (Lee et al., 1992; Levine, 2011; Raschke et al., 1999). The deletion of the GAGA element upstream of the *Hsp70* core promoter or the deletion of the intron enhancer and *C* regions of the *Ig kappa* gene abolished the occupancy of Pol II at their respective promoters. However, whether there are enhancer-associated factors that are more generally required for the establishment of Pol II at developmental genes was not known. Here, I show that the elongation factor Ell3 preferentially binds to enhancers, mediates the promoter-proximal occupancy of RNA Pol II at many of the developmentally regulated genes in mouse embryonic stem cells, and is required for their future activation during

stem cell specification (Figures 5.21A-C) (Lin et al., 2013). The Ell3-mediated enhancer function in promoter-proximal occupancy by Pol II requires the cohesin complex, revealing a novel step in the establishment of the "paused Pol II" state that is pervasive in ES cells and the regulation of gene activation during early embryo development.

5.3.1. A Model for the enhancer-associated Ell3 in the coordinated induction of transcription by SEC

Cohesin and mediator complexes were initially shown to be involved in the enhancer-promoter communication at active genes (Newman and Young, 2010). Interestingly, cohesin and mediator are also present at Ell3-bound "inactive" enhancers in ES cells, such as the beta globin locus (Figure 5.21A). It is possible that during differentiation, lineage-specific transcription factors such as GATA1 and NFE2 (Bulger and Groudine, 2011; Deng and Blobel, 2010) can interact with enhancer and promoter elements to help bridge cohesin and mediator communication with Pol II at the promoter, a process that can be further stabilized by Ell3 (Figure 5.21B). Setting up looped chromosomal domains could form a constrained transcriptional state associated with bivalent mark of H3K4 and H3K27 methylation (Bernstein et al., 2006) in a progenitor cell before full transcriptional activation. In addition, we find that many of the genes showing reduced paused Pol II after Ell3 knockdown were also in our Group **B**, which lacked the bivalent mark. Group **B** genes may constitute a transition state between constrained expression of bivalently marked genes and activated transcription (Cui et al., 2009; Zhang et al., 2012).

The presence of Ell3 could be particularly important at genes requiring an Ell2 version of SEC to release poised Pol II through phosphorylation of DSIF/NELF and the Pol II CTD (Figure 5.21C) (Lin et al., 2011; Luo et al., 2012b). Upon differentiation, Ell2

within SEC is recruited to the genes with Ell3-occupied enhancers and is required for their activation. This can be attributed to the function of Ell2 within SEC in releasing paused Pol II (Lin et al., 2011). Therefore, in certain circumstances, the ability of Ell2 and Ell3 to form similar and dynamic complexes could underlie a mechanism for the transition from Ell3's presence at poised enhancers in ES cells to Ell2's role in the release of paused Pol II during gene activation. Enhancer-promoter interactions could facilitate local assembly and/or recruitment of functional SEC complexes for rapid, but regulated, gene activation. This model is supported by observations that: 1) Ell3 binding to enhancers is required for the recruitment of SEC during differentiation; 2) Mediator occupies Ell3-bound enhancers; and (3) the MED26 subunit is required for the recruitment of SEC to the *HSP70* and *MYC* genes (Kagey et al., 2010; Takahashi et al., 2011).

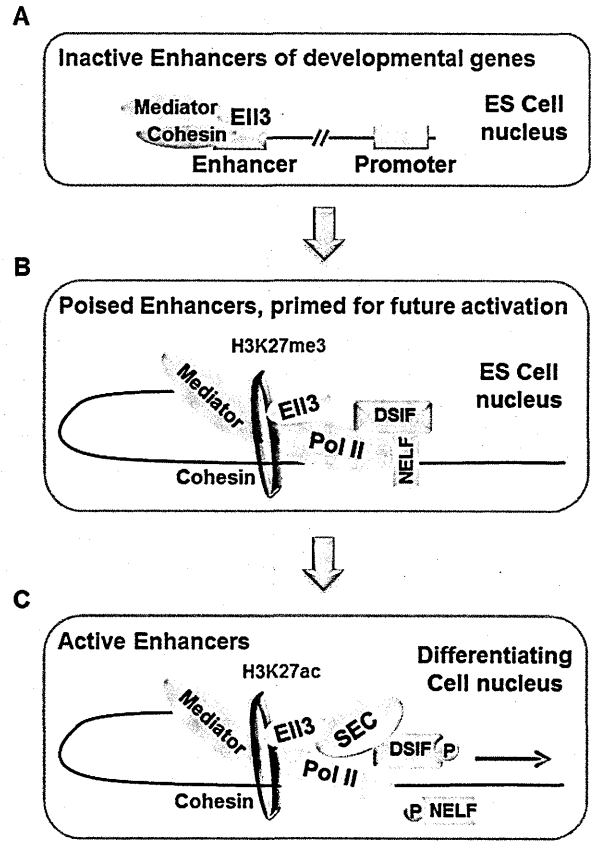


Figure 5.21. Model for the enhancer-associated elongation factor Ell3 in gene activation
A model for the enhancer-associated Ell3 in coordinated transcriptional induction by SEC. H3K4me1, p300, Mediator, and cohesin can be found with Ell3 at inactive, poised, and active enhancers. At inactive enhancers, Ell3 is prebound with Mediator and cohesin, but Pol II is not found at the promoter. In the poised state, a subset of developmental regulators is in a constrained state of expression, with both H3K4me3 and H3K27me3 at the promoter. Pol II's presence at these promoters depends on the interactions between cohesin, Mediator, and Ell3. Bottom panel, upon receiving the proper activating signals, SEC is recruited and stabilized at the promoter region through interaction with Mediator and Ell3. SEC phosphorylates RNA Pol II CTD, Spt5, and Nelf, thus resulting in the release of Pol II and gene activation.

5.3.2. Ell3 as a candidate for priming future gene activation

The activation of the zygotic genome, leading to the control of development by both the paternal and the maternal genomes, is a key event during the maternal to zygotic transition (MZT) following fertilization. Recent studies have identified the

zinc finger protein, Zelda, as a factor that marks the promoter and enhancer regions of both active and inactive genes in the *Drosophila* early zygotic genome (Harrison et al., 2011; Liang et al., 2008; Nien et al., 2011). The association of Zelda with inactive genes is required for their future activation, indicating an essential role of Zelda as a "pioneer transcription factor" (Harrison et al., 2011). However, how Zelda binding at the promoter and enhancer regions of inactive genes regulates their future expression and how the cofactors working together with Zelda to activate the zygotic genome remain largely unknown.

In this study, I found that mammalian Ell3 not only binds to the enhancer regions of active genes, but also marks the enhancers of inactive genes in mouse embryonic stem cells, many of which are lineage-specific genes, such as beta-globins, *Gsc* and *T*. The marking of Ell3 at the inactive enhancers of mammalian cells is required for the future activation of their associated genes, analogous to what has been reported for Zelda in *Drosophila* embryogenesis. There is no Zelda homolog known outside of arthropods (Liang et al., 2008), and the mechanism by which Ell3 is so generally recruited to enhancers of varying chromatin states and transcriptional activities is currently unknown. Instead of a single Zelda in mammals, there could conceivably be a large family of zinc finger proteins that can recognize various enhancer sequences and help recruit Ell3.

Intriguingly, both Ell3 and Pol II seem to be associated in the nuclei of mouse sperm, but do not co-localize, which might be explained if Ell3 was occupying inactive enhancers and Pol II was present at TSS regions, similar to what we observe by ChIP-seq analyses in the ES cell state. These data suggest that Ell3 might serve as an epigenetic marker in germ cells by bookmarking the inactive enhancers of genes

for future activation in the embryo. ChIP-sequencing analyses of Ell3 and Pol II in sperm and oocytes will be required to test this model and could reveal interesting information on mechanisms of epigenetic inheritance.

5.3.3. A potential role of Ell3 in cancer pathogenesis

Many of the SEC components are among the most frequent MLL (myeloid/lymphoid or mixed-lineage leukemia) translocation partners found in human AML and ALL leukemia patients (Mohan et al., 2010b; Smith et al., 2011a). SEC is broadly recruited to MLL chimera target genes in both MLL-SEC and MLL-non-SEC-translocated leukemic cells (Lin et al., 2010; Yokoyama et al., 2010). Interestingly, many of the commonly and highly mis-regulated genes by different MLL chimeras including the *Runx1*, *Ebf1*, *Cdk6*, *Meis1/2,* and *Hoxa* cluster genes (Dawson et al., 2011; Krivtsov and Armstrong, 2007; Lin et al., 2010) are occupied by Ell3 on their enhancers in the ES state. It is likely that the MLL chimeras bypass the tight regulation of these ELL3-associated genes in hematopoietic cells, contributing to leukemogenesis. Therefore, investigating the extent to which Ell3 functions in other stem/progenitor cells could have implications in SEC function in other developmental pathways and the mis-regulation of SEC in disease.

Chapter 6. Discussion

Transcriptional elongation has now emerged as a common and rate-limiting regulatory step in the process of transcribing genetic information from DNA into functional RNA molecules in metazoans, as suggested by the genome-wide data that Pol II is paused proximally to the promoters of many developmentally regulated genes, associating with a short capped transcript (Levine, 2011; Muse et al., 2007; Nechaev et al., 2010; Zeitlinger et al., 2007). Release of Pol II from the pausing site into productive elongation is triggered by the coordinated action of the positive elongation factor, P-TEFb, and negative elongation factors, NELF and DSIF (Gilmour, 2009; Peterlin and Price, 2006).

In the present studies, the Super Elongation Complex (SEC) was identified through biochemical purification and proteomic analyses. SEC contains ELL family members (ELL1-3), MLL translocation partners (AFF1, AFF4, ENL, and AF9), and the Pol II elongation factor, P-TEFb (Lin et al., 2010) (Figure 1A). It was found here that as one of the most active P-TEFb-containing complexes, SEC plays an important role in the regulated release of paused Pol II and gene activation in a paused Pol II-dependent and independent manner (Lin et al., 2011; Lin et al., 2010; Luo et al., 2012b; Smith et al., 2011b). Furthermore, the requirement of MLL-SEC for the aberrant activation of the MLL chimera target genes suggests a critical role for abnormal transcription elongation during leukemogenesis (Lin et al., 2011; Lin et al., 2010; Mohan et al., 2010b; Smith et al., 2011a). In this section, I will generally discuss the current understanding of the SEC family complexes in rapid transcriptional activation of genes during development. Also, I will cover the emerging roles of SEC in cancer pathogenesis.

6.1. Different P-TEFb-containing complexes

6.1.1. The SEC family of P-TEFb-containing complexes

AFF4, the central protein tethering other components in SEC, belongs to the AFF family in mammals, which also includes AFF1, AFF2, and AFF3. All of the members in this family are delineated by conserved N- and C-terminal homology regions, a transactivation domain rich in serine, and an AF4/LAF4/FMR2 ("ALF") homology domain (Bitoun and Davies, 2005). Besides SEC, SEC-like 2 (SEC-L2) and SEC-like 3 (SEC-L3) were also identified through biochemically purifying AFF2 and AFF3, respectively (Luo et al; 2012). Similar to SEC, SEC-like complexes also contain ENL/AF9 and the kinase module P-TEFb (Figure 6.1). However, the presence of ELL proteins in SEC-L2/3 still remains elusive, since ELLs are not observed by the MudPIT analyses. SEC-like complexes are also active P-TEFb complexes as shown by *in vitro* Pol II CTD kinase assays(Luo et al., 2012a). Therefore, SEC, in a broader sense, refers to a series of active P-TEFb-containing complexes using different AFF family members as a scaffold in the absence or presence of ELL1-3. In addition, the two YEATS (Ynl107, ENL, AF9, and TFIIF small subunit) family members, ENL and AF9, also reside in separate SECs. Many versions of SEC generated from different combinations of these components would expand the regulatory ability of the SECs.

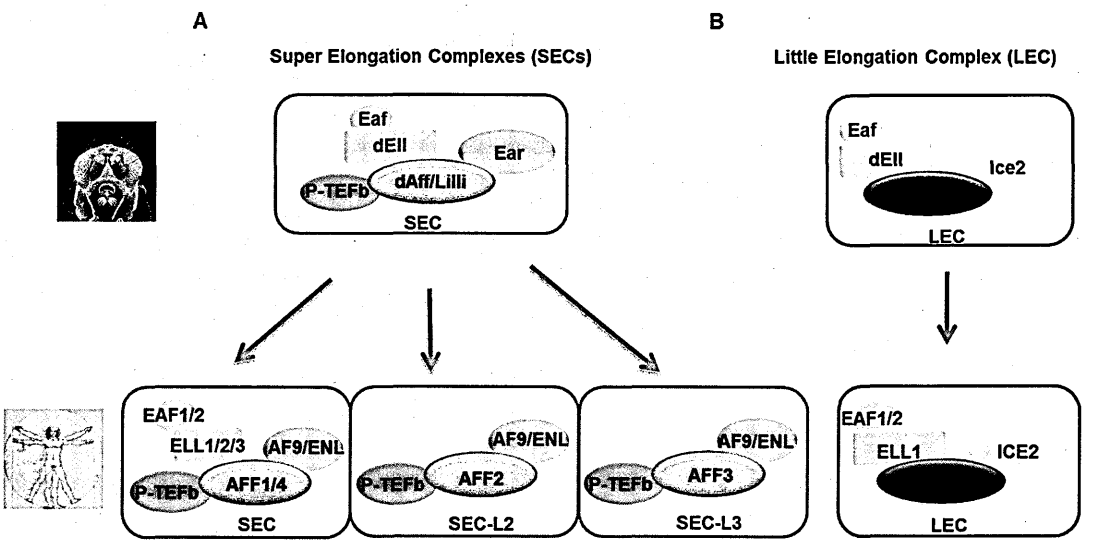


Figure 6.1. Distinct classes of genes are regulated by SEC and LEC families from *Drosophila* to human.

(A) In mammals, the four members of the AFF family (which comprises AFF1, AFF2, AFF3, and AFF4) and the three members of the ELL family of RNA polymerase II elongation factors (which comprises ELL1, ELL2, and ELL3) can be found in SEC complexes with compositional and functional diversity. AFF2 and AFF3 are the central factors in the formation of the SEC-like complexes SEC-L2 and SEC-L3, respectively. Like SEC, SEC-L complexes also contain the positive transcription elongation factor P-TEFb and ENL or AF9. The presence of the ELL family members in these two complexes has not been detected in biochemical purifications of AFF2 and AFF3. SEC containing ELL2 and AFF4 is involved in the Transcriptional Elongation Checkpoint Control (TECC) of paused Pol II-dependent and -independent rapid transcriptional induction. SEC in *Drosophila* has a similar subunit composition to mammalian SEC, but there is only one homolog for each of the AFF, ELL, EAF 1 and 2 (ELL-associated factor 1 and 2), and AF9 or ENL families: Lilli, related to AFF1-4 proteins; dEll, related to ELL1-3; Eaf, related to EAF1 and EAF2; and Ear, ENL and AF9-related, respectively. (B) *Drosophila* LEC contains dEll, Ice 1 and 2 (Interacts with C-terminus of ELL 1 and 2) and Eaf, and regulates the transcription of Pol II-transcribed snRNA genes. Although there are three members of the ELL family in mammals, only ELL1 is found to affect snRNA gene transcription as part of a mammalian LEC complex. It has been demonstrated that the ELL-containing SEC is required for rapid transcriptional induction of genes transcribed by Pol II and that the ELL-containing LEC is required for Pol II transcribed snRNA-encoding genes (Lin et al., 2011; Smith et al., 2011b).

6.1.2. Gene target specificities of the SEC and SEC-like complexes

The phosphorylation of the Pol II CTD, DSIF, and NELF by active P-TEFb are essential for the release of Pol II from the promoter-proximal pause sites (Peterlin and Price, 2006). Up to now, four different active P-TEFb-containing complexes have been

identified in mammals, including BRD4/P-TEFb, SEC, SEC-L2 and SEC-L3 (Table 6-1). In mammalian cells, both inactive and active forms of P-TEFb co-exist in equilibrium. Most of P-TEFb is insulated within the inactive 7SK snRNP-containing complex (Table 6-1). When the cellular need arises, P-TEFb is dissociated from the inactive pool and recruited to the chromatin to stimulate transcriptional elongation through interacting with Bromodomain-containing protein 4 (BRD4) or incorporating into SEC or SEC-like complexes.

It would be interesting to understand why in a variety of cellular contexts two or more of these active P-TEFb complexes coexist and the functional differences among them. Firstly, during both basal and activated transcription, the actions of P-TEFb are needed to allow paused Pol II to shift into the productive elongation stage. Inhibition of P-TEFb by the CDK9 inhibitor, flavopiridol, not only suppresses the global transcription, but also blocks the release of paused Pol II from the promoter-proximal region after several hours of treatment, suggesting the central role of P-TEFb in regulating transcription in the resting state (Rahl et al., 2010). Secondly, each of these active P-TEFb complexes could have its own specialty *in vivo* (Luo et al., 2012a). For example, although possessing similar CTD kinase activities *in vitro*, SEC-L2 and -L3 are not recruited to the promoter of the *HSP70* gene upon stress and also not required for its proper induction, indicating that different P-TEFb-containing complexes may have their own functional target preference. This notion was supported by the genome-wide expression data showing that SEC, SEC-L2 and SEC-L3 regulate different subsets of genes *in vivo*, with SEC focusing on these immediate response genes. Furthermore, all of the AFFs are linked to human diseases. The unique roles of the AFFs in different diseases might be caused by the target specificities of the SECs.

Table 6-1. Different P-TEFb-containing complexes in mammals

7SK/P-TEFb	BRD4/P-TEFb	SEC	SEC-L2	SEC-L3	P-TEFb
7SK					7SK
hnRNP					hnRNP
Hexim1/2					Hexim1/2
LARP7					LARP7
MEPCE					MEPCE
	BRD4				BRD4
		AFF1/4			AFF1/4
			AFF2		AFF2
				AFF3	AFF3
		EAF1/2			EAF1/2
		ELL1-3			ELL1-3
		AF9/ENL	AF9/ENL	AF9/ENL	AF9/ENL
CDK9	CDK9	CDK9	CDK9	CDK9	CDK9
Cyclin T	Cyclin T	Cyclin T	Cyclin T	Cyclin T	Cyclin T

It should be noted here that in some cases different P-TEFb complexes could work together to fine tune the expression level of the same gene. One example is the above-mentioned HIV-1 LTR. Another example is the regulation of *MYC* gene expression by both SEC and BRD4 (Dawson et al., 2011; Delmore et al., 2011; Luo et al., 2012a; Zuber et al., 2011). Inhibition of BRD2/3/4 by inhibitor I-BET151 in leukemia cells not only reduces the binding of BRD2/3/4 to the chromatin, but also dislodges the SEC and PAF1 complex from the chromatin (Dawson et al., 2011).

6.2. SEC in rapid gene induction

Proper adaptation to acute stress such as nutrient or temperature variation and developmental cues requires efficient activation of immediate response genes, as most of these genes encode master regulators controlling the expression of a plethora of down-stream effectors, which will in turn alter the transcriptional network and physiological function of the cells to maximize the cellular adaptation capacity.

6.2.1. SEC in Pol II-dependent rapid transcriptional activation

In vitro studies showed that both BRD4/P-TEFb and SEC could phosphorylate Pol II CTD (He and Zhou, 2011; Luo et al., 2012a; Smith et al., 2011a). However, the full extent of the functional differences between the active versions of the P-TEFb-containing complexes is not yet known. Although BRD4 is recruited to the HIV-1 promoter, it only affects basal transcription (Jang et al., 2005; Yang et al., 2005). The functional evidence from the SEC studies shows that SEC is required for stress-induced *HSP70* gene expression and Tat-transactivated HIV-1 LTR transcription (He et al., 2010; Lin et al., 2010; Sobhian et al., 2010).

In mouse embryonic stem cells, SEC components peak at promoter-proximal regions and travel with Pol II along the gene body on highly transcribed genes (Lin et al., 2011). In general, rapid induction is now recognized as the main function of SEC: recent whole genome studies revealed that the loading of SEC onto many of the retinoic acids (RA) responsive genes, such as *Hoxa1*, *caudal type homeobox 1 (Cdx1),* and *activating transcription factor 3* (*Atf3*), bearing paused Pol II is required for their rapid induction in response to RA. It should be mentioned here that RA treatment also leads to the recruitment of Brd4 to the same SEC-occupied genes. However, depletion of Brd4 via RNAi has little effect on the activation of most of these genes, suggesting that SEC, but

not BRD4/P-TEFb, is one of the major effectors of rapid induction of gene expression. The co-recruitment, but differential requirement, of SEC and BRD4/P-TEFb implicates that within the transcription cycle each elongation factor has specialized roles, perhaps working together in some cellular contexts.

6.2.2. SEC in Pol II-independent rapid transcriptional activation

It has been suggested that promoter proximally paused Pol II creates an advantage for efficient gene activation, since the polymerase is already in the right place waiting for a proper signal(s) for departure. Based on this point, it has been postulated that promoter-proximally paused Pol II is the basis for the rapid transcriptional induction of immediate response genes, including heat shock genes, serum-inducible genes, and certain developmentally controlled genes. However, it is not clear whether paused Pol II is always a prerequisite for rapid induction. Furthermore, it has also been proposed that the presence of paused Pol II is required for the synchronous induction of developmentally regulated genes (Gilmour, 2009; Levine, 2011).

In mammalian cells, the first example of rapid transcriptional induction in the absence of paused Pol II was shown for the activation of the *Cyp26a1* gene in mouse embryonic stem cells (mESC) by RA (Lin et al., 2011). The *Cyp26a1* gene encodes a member of the cytochrome P450 family of enzymes that tightly controls the level and distribution of RA through oxidative metabolism during embryogenesis (Tang and Gudas, 2011). In the undifferentiated mouse ES cells, the *Cyp26a1* gene is inactive and is covered by trimethylated H3K27, a mark of transcriptionally silenced chromatin. Furthermore, the *Cyp26a1* gene does not contain any detectable paused Pol II on its promoter. However, *Cyp26a1* is much more rapidly induced in the presence of RA than other paused Pol II bearing RA target genes (Lin et al., 2011) and Pol II and SEC are recruited to the *Cyp26a1*

gene within a few hours following RA treatment. Disruption of SEC, through either shRNA-mediated Ell2 depletion or inhibition of the kinase module Cdk9 by flavopiridol, greatly reduces the RA-activated *Cyp26a1* gene expression. This case indicates that paused Pol II is not strictly required for the rapid induction of transcription.

Moreover, in *Drosophila,* reducing promoter pausing of Pol II at the *Hsp70* locus by knocking down NELF does not slow the rate of gene activation after heat shock (Ghosh et al., 2011; Gilchrist et al., 2008). Based on this and other observations (Hah et al., 2011), paused Pol II might not be a prerequisite for rapid transcriptional induction. It could instead be that genes having paused Pol II in their promoter-proximal regions could respond to an inducing signal in a more synchronous and dynamic manner in the first several rounds of transcription, with the stochastic process of *de novo* Pol II recruitment having already been completed. Such a model was proposed to explain the association of paused Pol II with synchronous transcriptional induction in developing *Drosophila* embryos (Boettiger and Levine, 2009).

6.2.3. The potential roles of ELLs in transcriptional initiation control

The *in vitro* biochemistry elongation assays have shown that ELLs can enhance the elongation activity of Pol II by reducing transient pausing, which is further supported by the requirement of ELL2 for the release of paused Pol II from the *Hoxa1* promoter. However, the studies in mESC also demonstrated that the loss of SEC reduces both the proximal and elongating Pol II on the *Cyp26a1* gene, but has no significant effect on the paused Pol II on the *Hoxa1* gene (Lin et al., 2011). This phenomenon opens the possibility that, apart from participating in Pol II elongation, SEC and its component(s) may also help stabilize the rapidly assembled PIC on the *Cyp26a1* gene or those immediate response

genes without preloaded Pol II. This finding suggests the diverse mechanisms of SEC in regulating paused Pol II dependent and independent rapid gene activation.

6.3. Mechanisms for SEC recruitment

Translocations of any of the SEC subunits to the N-terminal DNA-binding domain of MLL improperly stabilizes SEC to MLL target genes, including *HOXA9* and *HOXA10*, leading to deviant gene expression and to aggressive acute leukemia (Lin et al., 2010; Smith et al., 2011a). Furthermore, the HIV-1 trans-activator of transcription (Tat) recruits SEC to activate HIV-1 gene expression (He et al., 2010; Lin et al., 2010; Mohan et al., 2010b; Sobhian et al., 2010). These findings provoke thoughts of how SEC is normally recruited and functions on its diverse target genes.

6.3.1. Recruitment of SEC by DNA-specific factors

SEC could initially be recruited to specific genomic loci by sequence-specific factors, or after specific cellular events such as juxtaposition to MLL via chromosomal translocation. DNA-binding factors, whose activities can be triggered by signaling events, may recruit SEC to a specific set of genes whose activation will allow the cell fate transition. For example, upon RA-induced differentiation of mESC into a neuronal lineage, SEC was found to bind to RA receptor (RAR) target genes, including *Hoxa1*, *Cdx1*, *Cyp26a1*, and to regulate their expression and initiate a neuronal lineage (Lin et al., 2011). In this case, it is very likely that RAR, activated by retinoid signaling, is the DNA-binding factor that helps recruit SEC to its target genes to promote transcription. Subsequently, the DNA-binding factors that deliver SEC to specific loci may boost the activity of SEC locally by stabilizing its components or the architecture of the whole complex.

In the case of MLL-SEC translocations, the fusion of AFF1, AFF4, or ELL1 with MLL may stabilize each other and increase the residency time of SEC with the *HOX* loci, leading to increased transcription (Figure 6.2A). In the case of Tat-mediated transactivation, Tat not only recruits SEC, but also prevents the polyubiquitination and the proteasomal degradation of the SEC components (Bursen et al., 2004; He et al., 2010), thereby, increasing the local concentration of fully functional SEC (Figure 6.2B).

6.3.2. DNA-specific factors independent recruitment of SEC

It should be noted that SEC exists and functions independently of MLL or Tat. Recent genome-wide data substantiates this point by using high-throughput ChIP-sequencing to precisely map the SEC components throughout the mESC genome. In the cellular context without MLL translocation or HIV-1 infection, many genes such as the housekeeping histone genes, are co-occupied by multiple SEC components at high confidence levels (Lin et al., 2011). The question raised here is, under normal physiological conditions, how is SEC recruited to its gene targets? Proteomic analyses identified an interaction between the SEC components ELL-associated factor 1/2 (EAF1/2) and the N-terminal TFIIS-like domain of the transcriptional coactivator, the Mediator Complex subunit 26 (MED26) (Takahashi et al., 2011), suggesting a mechanism for the recruitment of SEC to genes. This notion was validated by manual ChIP assays on selected genes demonstrating that after MED26 depletion, the occupancy of the SEC components was decreased throughout the whole transcribed regions of the *MYC* and *HSP70* genes. In line with this observation, the knockdown of MED26 also attenuates the signal of Pol II Ser2 phosphorylation on the two genes, indicating that MED26 is required for the proper function of SEC on these loci. However, others have proposed modes of recruitment of SEC that depend on either BRD3/4 complexes (Dawson et al., 2011) or

interactions between AF9/ENL and the RNA Pol II-associated factor complex (PAFc), which promotes mature transcription in multiple ways, including chromatin remodeling, elongation, and polyadenylation (He et al., 2011) (Figure 6.2C). Therefore, genome-wide profiling of SEC components before and after depletion of the various SEC recruiters is essential in determining the relative dependency of these factors for SEC recruitment to specific gene targets.

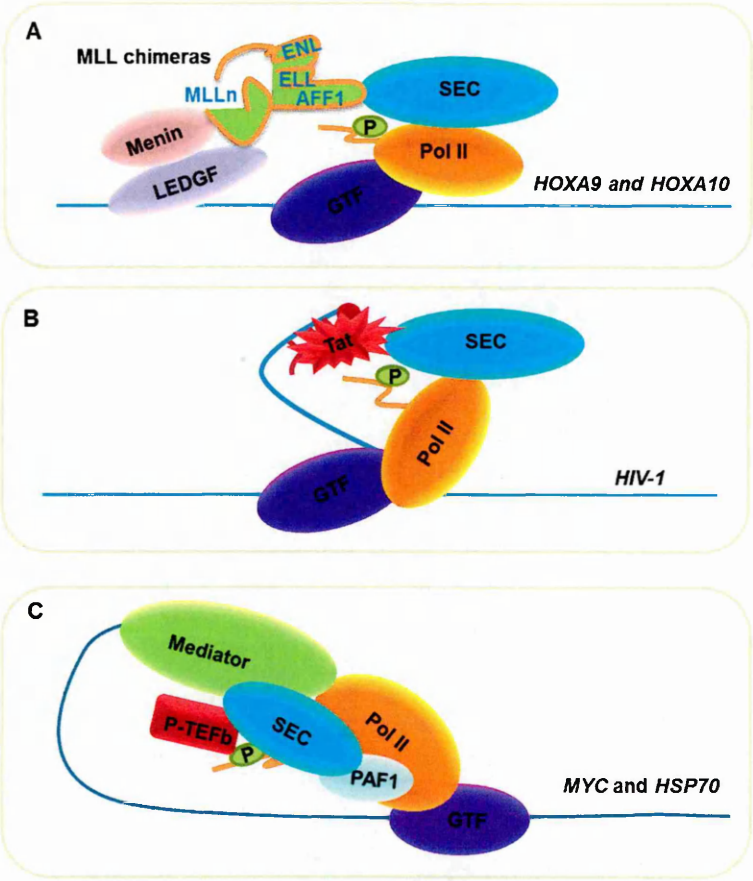


Figure 6.2. Diverse recruitment mechanisms of SEC in disease.

(A) In normal hematopoiesis, genes such as *HOXA9* and *HOXA10* are under strict transcriptional control. Chromosomal translocations that result in SEC subunits such as ENL, ELL, or AFF1 being fused to the N terminus of MLL (MLLn) promote the localization of SEC to *HOXA9* and *HOXA10* and also stabilize MLL-SEC locally, leading to misregulation of the TECC step and premature release of the paused Pol II on these genes without appropriate checkpoints. This misregulation of paused Pol II release has been proposed to be the molecular reason why translocations of MLL into so many unrelated genes results in leukemic pathogenesis. Menin and Lens Epithelial-Derived Growth Factor (LEDGF) are MLLn interactors, which are responsible for the recruitment of MLL chimeras to chromatin. GTF, general transcription factors. (B) During active HIV-1 infection, the viral transactivator of transcription, the Tat protein, recruits SEC to the HIV-1 long terminal repeat (LTR) to activate expression of the provirus in host cells. (C) In wild-type cells, it has been proposed that AFF4-containing SEC is recruited to its target genes (for example, *MYC* or *HSP70*) by the Mediator complex subunit, MED26, and the PAF1 complex to regulate their expression. Depletion of the bromodomain-containing protein, BRD4, another interactor of the positive transcription elongation factor, P-TEFb, also affects *MYC* expression (see Figure 6.3).

6.4. Target specificities of different elongation factors

As discussed in the Section 6.1.2, different active P-TEFb-containing complexes have their own target specificities *in vivo*. Furthermore, the function of Ell1 in snRNA gene transcription, Ell2 in rapid gene activation, and Ell3 marking at enhancers for setting up paused Pol II extend this principal: every elongation complex could have its own target. LEC contains ELL1, Interact with C-terminus ELL 1 (ICE1) and 2 (ICE2), in both *Drosophila* and mammals (Figure 6.1)(Smith et al., 2011b). In *Drosophila*, there is only one ELL protein, dELL. The dELL ChIP-seq combining with the total RNA-seq upon dELL RNAi showed that *Drosophila* dELL predominately regulates the transcription of Pol II-transcribed snRNA genes. However, there are three ELL paralogs in mammals. By taking the same strategy, I found that mammalian ELL1, instead of ELL2 and ELL3, is predominately involved in the regulation of snRNA genes. This finding suggested that different ELL proteins could function in different classes of genes and the specialization of LEC for snRNA genes and SEC for mRNA genes.

6.5. Transcription Elongation Checkpoint Control (TECC) and SEC in cancer

Transcription elongation checkpoint control (TECC), which is referred to as the regulation of Pol II set up and release from the paused state, is crucial for the regulation of gene expression during development and its misregulation can result in the onset of human diseases including cancer. As discussed above, the translocation of MLL to any of the SEC subunits can cause re-localization of SEC to the MLL target genes, resulting in an aberrant transcriptional elongation checkpoint defect and eventual leukemogenesis. In support of this model, I have demonstrated that AFF4, the platform for other SEC subunits association, is required for the complex stability, and that the knockdown of AFF4 in SEC

in leukemia cell lines reduces the expression of *HOXA9* and *HOXA10*, which are known targets of MLL.

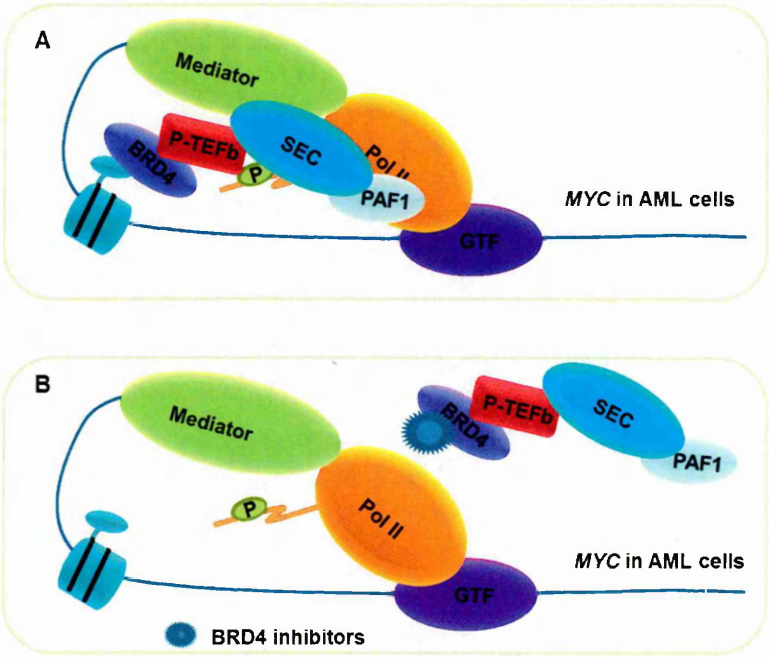


Figure 6.3. SEC in cancer progression.

(A) The expression of the *MYC* gene is regulated at the level of transcriptional elongation by paused Pol II through a mechanism known as the Transcriptional Elongation Checkpoint Control (TECC). AF4/FMR2 family member 4 (AFF4) occupies the entire transcription unit of the *MYC* gene and is required for its expression. Super Elongation Complex (SEC) containing AFF4 is recruited to the *MYC* locus by the Mediator complex subunit, MED26, and the PAF1 complex, which are both RNA polymerase II transcriptional coactivators (Takahashi et al., 2011). Reducing the level of MED26 or PAF1 by RNAi leads to a failure in the proper loading of the SEC components AFF4 and Cyclin-dependent kinase 9 (CDK9) to the *MYC* gene. Depletion of BRD4, another interactor of the positive elongation factor, P-TEFb, also affects *MYC* expression (Yang et al., 2008; Zuber et al., 2011). These findings raise the question of how SEC and BRD4 coordinate the regulation of *MYC* gene expression.
(B) In cells isolated from patients with acute myeloid leukemia (AML), BRD4 recruits the SEC and PAF1 complex to the *MYC* loci (Dawson et al., 2011). Inhibition of BRD4 by the small molecule I-BET151 leads to the dissociation of the SEC and the PAF1 complex from chromatin and results in the down-regulation of *MYC* gene expression (see the figure). The role of BDR4 in acute lymphoblastic leukemia (ALL) is not clear since these cells were insensitive to BRD4 inhibitors. Recently, it has been shown that SEC is required for the expression of the *MYC* gene in both AML and ALL (Luo et al., 2012a). Therefore, SEC could be a potential target for the treatment of the cancers with *MYC* gene overexpression.

Besides this, the finding that the cancer master gene *MYC* is a direct target of SEC implies that SEC could also function in cancer development and progression (Figure 6.3A). *MYC* is one of a few well-characterized genes which are mainly regulated at the level of transcription elongation by promoter-proximal paused Pol II. Inhibition of BRD4 by the small molecule JQ1 is able to efficiently arrest AML cell proliferation through reducing the MYC expression (Figure 6.3B). However, ALL growth is insensitive to BRD4 inhibition. Different from BRD4, the requirement of AFF4 on MYC expression was observed in both AML and ALL cells, suggesting that AFF4 could represent a broader spectrum target than BRD4 for therapeutic interventions in leukemia. Another target of SEC is the *ADAM metallopeptidase with thrombospondin type 1 motif, 1* (*ADAMTS1*) gene, which encodes a matrix degrading proteinase, establishing a permissive stromal microenvironment for tumor cell growth and migration. Therefore, SEC could function in cancer in three aspects: pathogenesis, progression, and metastasis. This positions SEC or its central factor, AFF4, as a broad-spectrum target for therapeutic interventions in leukemia and other cancers.

Chapter 7. Future Work

SEC family members have diverse and far-reaching effects on gene expression during normal development and during disease pathogenesis. The identification of the upstream signals that control the activity or stability of the SECs and deliver different SECs to different loci throughout the genome could help explain the differential recruitment of the SECs or SEC-related complexes. In order to delineate the pathways by which the SEC family promotes transcription, the identification of additional in vivo substrates for these P-TEFb-containing complexes should be considered, as Pol II, NELF, and DSIF might just be a fraction of proteins phosphorylated by P-TEFb within SECs. Furthermore, it will be of interest to determine the gene targets of the different SECs during development or under diverse conditions. This would also hopefully define the gene target selectivity for different Pol II elongation complexes and provide insights into how these complexes work together to satisfy the developmental needs. Given the success of the development of BRD4 inhibitors that function to disrupt BRD4 recruitment, and to reduce SEC-dependent MYC expression, it will be worth trying to develop inhibitors that more directly interfere with SEC family organization or stability. These could be used both to interrogate the normal function of the SEC family of complexes and as potential cancer therapeutics.

References

Abu-Abed, S, Dolle, P, Metzger, D, Beckett, B, Chambon, P, Petkovich, M (2001) The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. Genes Dev *15*, 226-240.

Agarwal, SK, Guru, SC, Heppner, C, Erdos, MR, Collins, RM, Park, SY, Saggar, S, Chandrasekharappa, SC, Collins, FS, Spiegel, AM, Marx, SJ, Burns, AL (1999) Menin interacts with the AP1 transcription factor JunD and represses JunD-activated transcription. Cell *96*, 143-152.

Ahn, SH, Kim, M, Buratowski, S (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. Mol Cell *13*, 67-76.

Aida, M, Chen, Y, Nakajima, K, Yamaguchi, Y, Wada, T, Handa, H (2006) Transcriptional pausing caused by NELF plays a dual role in regulating immediate-early expression of the junB gene. Mol Cell Biol *26*, 6094-6104.

Albert, A, Lavoie, S, Vincent, M (1999) A hyperphosphorylated form of RNA polymerase II is the major interphase antigen of the phosphoprotein antibody MPM-2 and interacts with the peptidyl-prolyl isomerase Pin1. J Cell Sci *112 (Pt 15)*, 2493-2500.

Alexander, T, Nolte, C, Krumlauf, R (2009) Hox genes and segmentation of the hindbrain and axial skeleton. Annu Rev Cell Dev Biol *25*, 431-456.

Andrulis, ED, Guzman, E, Doring, P, Werner, J, Lis, JT (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*, 2635-2649.

Andrulis, ED, Werner, J, Nazarian, A, Erdjument-Bromage, H, Tempst, P, Lis, JT (2002) The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. Nature *420*, 837-841.

Ardehali, MB, Yao, J, Adelman, K, Fuda, NJ, Petesch, SJ, Webb, WW, Lis, JT (2009) Spt6 enhances the elongation rate of RNA polymerase II in vivo. EMBO J *28*, 1067-1077.

Arigo, JT, Eyler, DE, Carroll, KL, Corden, JL (2006) Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. Mol Cell *23*, 841-851.

Armstrong, SA, Staunton, JE, Silverman, LB, Pieters, R, den Boer, ML, Minden, MD, Sallan, SE, Lander, ES, Golub, TR, Korsmeyer, SJ (2002) MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet *30*, 41-47.

Aso, T, Lane, WS, Conaway, JW, Conaway, RC (1995) Elongin (SIII): a multisubunit regulator of elongation by RNA polymerase II. Science *269*, 1439-1443.

Axel, R, Cedar, H, Felsenfeld, G (1973) Synthesis of globin ribonucleic acid from duck-reticulocyte chromatin in vitro. Proc Natl Acad Sci U S A *70*, 2029-2032.

Ayton, PM, Cleary, ML (2003) Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9. Genes Dev *17*, 2298-2307.

Bach, C, Mueller, D, Buhl, S, Garcia-Cuellar, MP, Slany, RK (2009) Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2. Oncogene *28*, 815-823.

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

Banks, CA, Kong, SE, Spahr, H, Florens, L, Martin-Brown, S, Washburn, MP, Conaway, JW, Mushegian, A, Conaway, RC (2007) Identification and Characterization of a Schizosaccharomyces pombe RNA Polymerase II Elongation Factor with Similarity to the Metazoan Transcription Factor ELL. J Biol Chem *282*, 5761-5769.

Barboric, M, Lenasi, T, Chen, H, Johansen, EB, Guo, S, Peterlin, BM (2009) 7SK snRNP/P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development. Proc Natl Acad Sci U S A *106*, 7798-7803.

Bartkowiak, B, Liu, P, Phatnani, HP, Fuda, NJ, Cooper, JJ, Price, DH, Adelman, K, Lis, JT, Greenleaf, AL (2010) CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev *24*, 2303-2316.

Bentley, D (2002) The mRNA assembly line: transcription and processing machines in the same factory. Curr Opin Cell Biol *14*, 336-342.

Bernstein, BE, Mikkelsen, TS, Xie, X, Kamal, M, Huebert, DJ, Cuff, J, Fry, B, Meissner, A, Wernig, M, Plath, K, Jaenisch, R, Wagschal, A, Feil, R, Schreiber, SL, Lander, ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Beyer, AL, Osheim, YN (1988) Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev *2*, 754-765.

Birke, M, Schreiner, S, Garcia-Cuellar, MP, Mahr, K, Titgemeyer, F, Slany, RK (2002) The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. Nucleic Acids Res *30*, 958-965.

Bisgrove, DA, Mahmoudi, T, Henklein, P, Verdin, E (2007) Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription. Proc Natl Acad Sci U S A *104*, 13690-13695.

Bitoun, E, Davies, KE (2005) The robotic mouse: unravelling the function of AF4 in the cerebellum. Cerebellum *4*, 250-260.

Bitoun, E, Davies, KE (2009) The robotic mouse: understanding the role of AF4, a cofactor of transcriptional elongation and chromatin remodelling, in purkinje cell function. Cerebellum *8*, 175-183.

Bitoun, E, Oliver, PL, Davies, KE (2007) The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. Hum Mol Genet *16*, 92-106.

Blazek, D, Kohoutek, J, Bartholomeeusen, K, Johansen, E, Hulinkova, P, Luo, Z, Cimermancic, P, Ule, J, Peterlin, BM (2011) The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. Genes Dev *25*, 2158-2172.

Boehm, AK, Saunders, A, Werner, J, Lis, JT (2003) Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock. Mol Cell Biol *23*, 7628-7637.

Boettiger, AN, Levine, M (2009) Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science *325*, 471-473.

Bradsher, JN, Jackson, KW, Conaway, RC, Conaway, JW (1993a) RNA polymerase II transcription factor SIII. I. Identification, purification, and properties. J Biol Chem *268*, 25587-25593.

Bradsher, JN, Tan, S, McLaury, HJ, Conaway, JW, Conaway, RC (1993b) RNA polymerase II transcription factor SIII. II. Functional properties and role in RNA chain elongation. J Biol Chem *268*, 25594-25603.

Breen, TR, Harte, PJ (1993) Trithorax regulates multiple homeotic genes in the bithorax and Antennapedia complexes and exerts different tissue-specific, parasegment-specific and promoter-specific effects on each. Development *117*, 119-134.

Bres, V, Yoh, SM, Jones, KA (2008) The multi-tasking P-TEFb complex. Curr Opin Cell Biol *20*, 334-340.

Broeker, PL, Super, HG, Thirman, MJ, Pomykala, H, Yonebayashi, Y, Tanabe, S, Zeleznik-Le, N, Rowley, JD (1996) Distribution of 11q23 breakpoints within the MLL breakpoint cluster region in de novo acute leukemia and in treatment-related acute myeloid leukemia: correlation with scaffold attachment regions and topoisomerase II consensus binding sites. Blood *87*, 1912-1922.

Bulger, M, Groudine, M (2011) Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327-339.

Buratowski, S, Hahn, S, Guarente, L, Sharp, PA (1989) Five intermediate complexes in transcription initiation by RNA polymerase II. Cell *56*, 549-561.

Buratowski, S, Sharp, PA (1990) Transcription initiation complexes and upstream activation with RNA polymerase II lacking the C-terminal domain of the largest subunit. Mol Cell Biol *10*, 5562-5564.

Bursen, A, Moritz, S, Gaussmann, A, Dingermann, T, Marschalek, R (2004) Interaction of AF4 wild-type and AF4.MLL fusion protein with SIAH proteins: indication for t(4;11) pathobiology? Oncogene *23*, 6237-6249.

Carrillo Oesterreich, F, Bieberstein, N, Neugebauer, KM (2011) Pause locally, splice globally. Trends Cell Biol *21*, 328-335.

Carrillo Oesterreich, F, Preibisch, S, Neugebauer, KM (2010) Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. Mol Cell *40*, 571-581.

Castelli-Gair, JE, Garcia-Bellido, A (1990) Interactions of Polycomb and trithorax with cis regulatory regions of Ultrabithorax during the development of Drosophila melanogaster. EMBO J *9*, 4267-4275.

Chang, PY, Hom, RA, Musselman, CA, Zhu, L, Kuo, A, Gozani, O, Kutateladze, TG, Cleary, ML (2010) Binding of the MLL PHD3 finger to histone H3K4me3 is required for MLL-dependent gene transcription. J Mol Biol *400*, 137-144.

Chao, SH, Price, DH (2001) Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. J Biol Chem *276*, 31793-31799.

Cheng, B, Li, T, Rahl, PB, Adamson, TE, Loudas, NB, Guo, J, Varzavand, K, Cooper, JJ, Hu, X, Gnatt, A, Young, RA, Price, DH (2012) Functional association of Gdown1 with RNA polymerase II poised on human genes. Mol Cell *45*, 38-50.

Cheng, B, Price, DH (2007) Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. J Biol Chem *282*, 21901-21912.

Cheung, N, Chan, LC, Thompson, A, Cleary, ML, So, CW (2007) Protein arginine-methyltransferase-dependent oncogenesis. Nat Cell Biol *9*, 1208-1215.

Cho, YW, Hong, T, Hong, S, Guo, H, Yu, H, Kim, D, Guszczynski, T, Dressler, GR, Copeland, TD, Kalkum, M, Ge, K (2007) PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. J Biol Chem *282*, 20395-20406.

Chodosh, LA, Fire, A, Samuels, M, Sharp, PA (1989) 5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole inhibits transcription elongation by RNA polymerase II in vitro. J Biol Chem *264*, 2250-2257.

Cleary, ML (2009) Regulating the leukaemia stem cell. Best Pract Res Clin Haematol *22*, 483-487.

Collart, MA, Tourkine, N, Belin, D, Vassalli, P, Jeanteur, P, Blanchard, JM (1991) c-fos gene transcription in murine macrophages is modulated by a calcium-dependent block to elongation in intron 1. Mol Cell Biol *11*, 2826-2831.

Conaway, JW, Bond, MW, Conaway, RC (1987) An RNA polymerase II transcription system from rat liver. Purification of an essential component. J Biol Chem *262*, 8293-8297.

Conaway, JW, Conaway, RC (1989a) A multisubunit transcription factor essential for accurate initiation by RNA polymerase II. J Biol Chem *264*, 2357-2362.

Conaway, RC, Conaway, JW (1989b) An RNA polymerase II transcription factor has an associated DNA-dependent ATPase (dATPase) activity strongly stimulated by the TATA region of promoters. Proc Natl Acad Sci U S A *86*, 7356-7360.

Conaway, RC, Conaway, JW (1993) General initiation factors for RNA polymerase II. Annu Rev Biochem *62*, 161-190.

Corden, JL (1990) Tails of RNA polymerase II. Trends Biochem Sci *15*, 383-387.

Core, LJ, Waterfall, JJ, Lis, JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Cozzio, A, Passegue, E, Ayton, PM, Karsunky, H, Cleary, ML, Weissman, IL (2003) Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. Genes Dev *17*, 3029-3035.

Cramer, P (2004) Structure and function of RNA polymerase II. Adv Protein Chem *67*, 1-42.

Cramer, P, Armache, KJ, Baumli, S, Benkert, S, Brueckner, F, Buchen, C, Damsma, GE, Dengl, S, Geiger, SR, Jasiak, AJ, Jawhari, A, Jennebach, S, Kamenski, T, Kettenberger, H, Kuhn, CD, Lehmann, E, Leike, K, Sydow, JF, Vannini, A (2008) Structure of eukaryotic RNA polymerases. Annu Rev Biophys *37*, 337-352.

Cramer, P, Bushnell, DA, Kornberg, RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science *292*, 1863-1876.

Cremer, T, Cremer, M (2010) Chromosome territories. Cold Spring Harb Perspect Biol *2*, a003889.

Creyghton, MP, Cheng, AW, Welstead, GG, Kooistra, T, Carey, BW, Steine, EJ, Hanna, J, Lodato, MA, Frampton, GM, Sharp, PA, Boyer, LA, Young, RA, Jaenisch, R (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A *107*, 21931-21936.

Cui, K, Zang, C, Roh, TY, Schones, DE, Childs, RW, Peng, W, Zhao, K (2009) Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell *4*, 80-93.

D'Orso, I, Frankel, AD (2010) RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation. Nat Struct Mol Biol *17*, 815-821.

Dahmus, ME (1995) Phosphorylation of the C-terminal domain of RNA polymerase II. Biochim Biophys Acta *1261*, 171-182.

Davison, BL, Egly, JM, Mulvihill, ER, Chambon, P (1983) Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. Nature *301*, 680-686.

Dawson, MA, Prinjha, RK, Dittmann, A, Giotopoulos, G, Bantscheff, M, Chan, WI, Robson, SC, Chung, CW, Hopf, C, Savitski, MM, Huthmacher, C, Gudgin, E, Lugo, D, Beinke, S, Chapman, TD, Roberts, EJ, Soden, PE, Auger, KR, Mirguet, O, Doehner, K, Delwel, R, Burnett, AK, Jeffrey, P, Drewes, G, Lee, K, Huntly, BJ, Kouzarides, T (2011) Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. Nature *478*, 529-533.

Delmore, JE, Issa, GC, Lemieux, ME, Rahl, PB, Shi, J, Jacobs, HM, Kastritis, E, Gilpatrick, T, Paranal, RM, Qi, J, Chesi, M, Schinzel, AC, McKeown, MR, Heffernan, TP, Vakoc, CR, Bergsagel, PL, Ghobrial, IM, Richardson, PG, Young, RA, Hahn, WC, Anderson, KC, Kung, AL, Bradner, JE, Mitsiades, CS (2011) BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Cell *146*, 904-917.

Deng, W, Blobel, GA (2010) Do chromatin loops provide epigenetic gene expression states? Curr Opin Genet Dev *20*, 548-554.

DiMartino, JF, Miller, T, Ayton, PM, Landewe, T, Hess, JL, Cleary, ML, Shilatifard, A (2000) A carboxy-terminal domain of ELL is required and sufficient for immortalization of myeloid progenitors by MLL-ELL. Blood *96*, 3887-3893.

Dixon, JR, Selvaraj, S, Yue, F, Kim, A, Li, Y, Shen, Y, Hu, M, Liu, JS, Ren, B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Djabali, M, Selleri, L, Parry, P, Bower, M, Young, BD, Evans, GA (1992) A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. Nat Genet *2*, 113-118.

Dobson, CL, Warren, AJ, Pannell, R, Forster, A, Rabbitts, TH (2000) Tumorigenesis in mice with a fusion of the leukaemia oncogene Mll and the bacterial lacZ gene. EMBO J *19*, 843-851.

Donner, AJ, Ebmeier, CC, Taatjes, DJ, Espinosa, JM (2010) CDK8 is a positive regulator of transcriptional elongation within the serum response network. Nat Struct Mol Biol *17*, 194-201.

Dorsett, D (2011) Cohesin: genomic insights into controlling gene transcription and development. Current opinion in genetics & development *21*, 199-206.

Dou, Y, Milne, TA, Ruthenburg, AJ, Lee, S, Lee, JW, Verdine, GL, Allis, CD, Roeder, RG (2006) Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat Struct Mol Biol *13*, 713-719.

Duboule, D, Dolle, P (1989) The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. EMBO J *8*, 1497-1505.

Duester, G (2008) Retinoic acid synthesis and signaling during early organogenesis. Cell *134*, 921-931.

Dvir, A (2002) Promoter escape by RNA polymerase II. Biochim Biophys Acta *1577*, 208-223.

Dynlacht, BD, Hoey, T, Tjian, R (1991) Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell *66*, 563-576.

Egloff, S, O'Reilly, D, Chapman, RD, Taylor, A, Tanzhaus, K, Pitts, L, Eick, D, Murphy, S (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. Science *318*, 1777-1779.

Eilers, M, Eisenman, RN (2008) Myc's broad reach. Genes Dev *22*, 2755-2766.

Eissenberg, JC, Ma, J, Gerber, MA, Christensen, A, Kennison, JA, Shilatifard, A (2002) dELL is an essential RNA polymerase II elongation factor with a general role in development. Proc Natl Acad Sci U S A *99*, 9894-9899.

Eissenberg, JC, Shilatifard, A (2010) Histone H3 lysine 4 (H3K4) methylation in development and differentiation. Dev Biol *339*, 240-249.

Eissenberg, JC, Shilatifard, A, Dorokhov, N, Michener, DE (2007) Cdk9 is an essential kinase in Drosophila that is required for heat shock gene expression, histone methylation and elongation factor recruitment. Mol Genet Genomics *277*, 101-114.

Elmendorf, BJ, Shilatifard, A, Yan, Q, Conaway, JW, Conaway, RC (2001) Transcription factors TFIIF, ELL, and Elongin negatively regulate SII-induced nascent transcript cleavage by non-arrested RNA polymerase II elongation intermediates. J Biol Chem *276*, 23109-23114.

Engel, N, West, AG, Felsenfeld, G, Bartolomei, MS (2004) Antagonism between DNA hypermethylation and enhancer-blocking activity at the H19 DMD is uncovered by CpG mutations. Nat Genet *36*, 883-888.

Erfurth, F, Hemenway, CS, de Erkenez, AC, Domer, PH (2004) MLL fusion partners AF4 and AF9 interact at subnuclear foci. Leukemia *18*, 92-102.

Estable, MC, Naghavi, MH, Kato, H, Xiao, H, Qin, J, Vahlne, A, Roeder, RG (2002) MCEF, the newest member of the AF4 family of transcription factors involved in leukemia, is a positive transcription elongation factor-b-associated protein. J Biomed Sci *9*, 234-245.

Fanghanel, J, Fischer, G (2004) Insights into the catalytic mechanism of peptidyl prolyl cis/trans isomerases. Front Biosci *9*, 3453-3478.

Feaver, WJ, Svejstrup, JQ, Henry, NL, Kornberg, RD (1994) Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. Cell *79*, 1103-1109.

Finkel, JS, Chinchilla, K, Ursic, D, Culbertson, MR (2010) Sen1p performs two genetically separable functions in transcription and processing of U5 small nuclear RNA in Saccharomyces cerevisiae. Genetics *184*, 107-118.

Fire, A, Samuels, M, Sharp, PA (1984) Interactions between RNA polymerase II, factors, and template leading to accurate transcription. J Biol Chem *259*, 2509-2516.

Flores, O, Maldonado, E, Reinberg, D (1989) Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. J Biol Chem *264*, 8913-8921.

Fuda, NJ, Ardehali, MB, Lis, JT (2009) Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186-192.

Fujinaga, K, Irwin, D, Huang, Y, Taube, R, Kurosu, T, Peterlin, BM (2004) Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. Mol Cell Biol *24*, 787-795.

Gall, JG, Bellini, M, Wu, Z, Murphy, C (1999) Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. Mol Biol Cell *10*, 4385-4402.

Gautier, L, Cope, L, Bolstad, BM, Irizarry, RA (2004) affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics *20*, 307-315.

Gavalas, A, Trainor, P, Ariza-McNaughton, L, Krumlauf, R (2001) Synergy between Hoxa1 and Hoxb1: the relationship between arch patterning and the generation of cranial neural crest. Development *128*, 3017-3027.

Gerber, M, Ma, J, Dean, K, Eissenberg, JC, Shilatifard, A (2001) Drosophila ELL is associated with actively elongating RNA polymerase II on transcriptionally active sites in vivo. EMBO J *20*, 6104-6114.

Gerber, M, Tenney, K, Conaway, JW, Conaway, RC, Eissenberg, JC, Shilatifard, A (2005a) Regulation of heat shock gene expression by RNA polymerase II elongation factor, Elongin A. J Biol Chem *280*, 4017-4020.

Gerber, MA, Shilatifard, A, Eissenberg, JC (2005b) Mutational analysis of an RNA polymerase II elongation factor in Drosophila melanogaster. Mol Cell Biol *25*, 7803-7811.

Gerstein, MB, Lu, ZJ, Van Nostrand, EL, Cheng, C, Arshinoff, BI, Liu, T, Yip, KY, Robilotto, R, Rechtsteiner, A, Ikegami, K, Alves, P, Chateigner, A, Perry, M, Morris, M, Auerbach, RK, Feng, X, Leng, J, Vielle, A, Niu, W, Rhrissorrakrai, K, Agarwal, A, Alexander, RP, Barber, G, Brdlik, CM, Brennan, J, Brouillet, JJ, Carr, A, Cheung, MS, Clawson, H, Contrino, S, Dannenberg, LO, Dernburg, AF, Desai, A, Dick, L, Dose, AC, Du, J, Egelhofer, T, Ercan, S, Euskirchen, G, Ewing, B, Feingold, EA, Gassmann, R, Good, PJ, Green, P, Gullier, F, Gutwein, M, Guyer, MS, Habegger, L, Han, T, Henikoff, JG, Henz, SR, Hinrichs, A, Holster, H, Hyman, T, Iniguez, AL, Janette, J, Jensen, M, Kato, M, Kent, WJ, Kephart, E, Khivansara, V, Khurana, E, Kim, JK, Kolasinska-Zwierz, P, Lai, EC, Latorre, I, Leahey, A, Lewis, S, Lloyd, P, Lochovsky, L, Lowdon, RF, Lubling, Y, Lyne, R, MacCoss, M, Mackowiak, SD, Mangone, M, McKay, S, Mecenas, D, Merrihew, G, Miller, DM, 3rd, Muroyama, A, Murray, JI, Ooi, SL, Pham, H, Phippen, T, Preston, EA, Rajewsky, N, Ratsch, G, Rosenbaum, H, Rozowsky, J, Rutherford, K, Ruzanov, P, Sarov, M, Sasidharan, R, Sboner, A, Scheid, P, Segal, E, Shin, H, Shou, C, Slack, FJ, Slightam, C, Smith, R, Spencer, WC, Stinson, EO, Taing, S, Takasaki, T, Vafeados, D, Voronina, K, Wang, G, Washington, NL, Whittle, CM, Wu, B, Yan, KK, Zeller, G, Zha, Z, Zhong, M, Zhou, X, Ahringer, J, Strome, S, Gunsalus, KC, Micklem, G, Liu, XS, Reinke, V, Kim, SK, Hillier, LW, Henikoff, S, Piano, F, Snyder, M, Stein, L, Lieb, JD, Waterston, RH (2010) Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science *330*, 1775-1787.

Ghosh, SK, Missra, A, Gilmour, DS (2011) Negative elongation factor accelerates the rate at which heat shock genes are shut off by facilitating dissociation of heat shock factor. Mol Cell Biol *31*, 4232-4243.

Giardina, C, Perez-Riba, M, Lis, JT (1992) Promoter melting and TFIID complexes on Drosophila genes in vivo. Genes Dev *6*, 2190-2200.

Gilchrist, DA, Nechaev, S, Lee, C, Ghosh, SK, Collins, JB, Li, L, Gilmour, DS, Adelman, K (2008) NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes Dev *22*, 1921-1933.

Gilmour, DS (2009) Promoter proximal pausing on genes in metazoans. Chromosoma *118*, 1-10.

Gilmour, DS, Lis, JT (1986) RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Mol Cell Biol *6*, 3984-3989.

Gilmour, RS, Paul, J (1973) Tissue-specific transcription of the globin gene in isolated chromatin. Proc Natl Acad Sci U S A *70*, 3440-3442.

Gnatt, AL, Cramer, P, Fu, J, Bushnell, DA, Kornberg, RD (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. Science *292*, 1876-1882.

Goo, YH, Sohn, YC, Kim, DH, Kim, SW, Kang, MJ, Jung, DJ, Kwak, E, Barlev, NA, Berger, SL, Chow, VT, Roeder, RG, Azorsa, DO, Meltzer, PS, Suh, PG, Song, EJ, Lee, KJ, Lee, YC, Lee, JW (2003) Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. Mol Cell Biol *23*, 140-149.

Graham, A, Papalopulu, N, Krumlauf, R (1989) The murine and Drosophila homeobox gene complexes have common features of organization and expression. Cell *57*, 367-378.

Gu, Y, Nakamura, T, Alder, H, Prasad, R, Canaani, O, Cimino, G, Croce, CM, Canaani, E (1992) The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to Drosophila trithorax, to the AF-4 gene. Cell *71*, 701-708.

Guenther, MG, Lawton, LN, Rozovskaia, T, Frampton, GM, Levine, SS, Volkert, TL, Croce, CM, Nakamura, T, Canaani, E, Young, RA (2008) Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. Genes Dev *22*, 3403-3408.

Guenther, MG, Levine, SS, Boyer, LA, Jaenisch, R, Young, RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

Hadjur, S, Williams, LM, Ryan, NK, Cobb, BS, Sexton, T, Fraser, P, Fisher, AG, Merkenschlager, M (2009) Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature *460*, 410-413.

Haering, CH, Lowe, J, Hochwagen, A, Nasmyth, K (2002) Molecular architecture of SMC proteins and the yeast cohesin complex. Mol Cell *9*, 773-788.

Hagege, H, Klous, P, Braem, C, Splinter, E, Dekker, J, Cathala, G, de Laat, W, Forne, T (2007) Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat Protoc *2*, 1722-1733.

Hah, N, Danko, CG, Core, L, Waterfall, JJ, Siepel, A, Lis, JT, Kraus, WL (2011) A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell *145*, 622-634.

Haltiner, MM, Smale, ST, Tjian, R (1986) Two distinct promoter elements in the human rRNA gene identified by linker scanning mutagenesis. Mol Cell Biol *6*, 227-235.

Hanada, K, Song, CZ, Yamamoto, K, Yano, K, Maeda, Y, Yamaguchi, K, Muramatsu, M (1996) RNA polymerase I associated factor 53 binds to the nucleolar transcription factor UBF and functions in specific rDNA transcription. EMBO J *15*, 2217-2226.

Harreman, M, Taschner, M, Sigurdsson, S, Anindya, R, Reid, J, Somesh, B, Kong, SE, Banks, CA, Conaway, RC, Conaway, JW, Svejstrup, JQ (2009) Distinct ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation. Proc Natl Acad Sci U S A *106*, 20705-20710.

Harrison, MM, Li, XY, Kaplan, T, Botchan, MR, Eisen, MB (2011) Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. PLoS Genet *7*, e1002266.

Hawley, DK, Roeder, RG (1985) Separation and partial characterization of three functional steps in transcription initiation by human RNA polymerase II. J Biol Chem *260*, 8163-8172.

He, N, Chan, CK, Sobhian, B, Chou, S, Xue, Y, Liu, M, Alber, T, Benkirane, M, Zhou, Q (2011) Human Polymerase-Associated Factor complex (PAFc) connects the Super Elongation Complex (SEC) to RNA polymerase II on chromatin. Proc Natl Acad Sci U S A *108*, E636-645.

He, N, Liu, M, Hsu, J, Xue, Y, Chou, S, Burlingame, A, Krogan, NJ, Alber, T, Zhou, Q (2010) HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. Mol Cell *38*, 428-438.

He, N, Zhou, Q (2011) New Insights into the Control of HIV-1 Transcription: When Tat Meets the 7SK snRNP and Super Elongation Complex (SEC). J Neuroimmune Pharmacol *6*, 260-268.

Heintzman, ND, Hon, GC, Hawkins, RD, Kheradpour, P, Stark, A, Harp, LF, Ye, Z, Lee, LK, Stuart, RK, Ching, CW, Ching, KA, Antosiewicz-Bourget, JE, Liu, H, Zhang, X, Green, RD, Lobanenkov, VV, Stewart, R, Thomson, JA, Crawford, GE, Kellis, M, Ren, B (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.

Hengartner, CJ, Myer, VE, Liao, SM, Wilson, CJ, Koh, SS, Young, RA (1998) Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. Mol Cell *2*, 43-53.

Ho, CK, Shuman, S (1999) Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. Mol Cell *3*, 405-411.

Ho, CK, Sriskanda, V, McCracken, S, Bentley, D, Schwer, B, Shuman, S (1998) The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem *273*, 9577-9585.

Holstege, FC, van der Vliet, PC, Timmers, HT (1996) Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. EMBO J *15*, 1666-1677.

Hou, C, Dale, R, Dean, A (2010) Cell type specificity of chromatin organization mediated by CTCF and cohesin. Proc Natl Acad Sci U S A *107*, 3651-3656.

Hsieh, JJ, Ernst, P, Erdjument-Bromage, H, Tempst, P, Korsmeyer, SJ (2003) Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. Mol Cell Biol *23*, 186-194.

Hsin, JP, Sheth, A, Manley, JL (2011) RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. Science *334*, 683-686.

Hsu, LM (2002) Promoter clearance and escape in prokaryotes. Biochim Biophys Acta *1577*, 191-207.

Huang da, W, Sherman, BT, Lempicki, RA (2009a) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.

Huang da, W, Sherman, BT, Lempicki, RA (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Huang, S, Litt, M, Felsenfeld, G (2005) Methylation of histone H4 by arginine methyltransferase PRMT1 is essential in vivo for many subsequent histone modifications. Genes Dev *19*, 1885-1893.

Hughes, CM, Rozenblatt-Rosen, O, Milne, TA, Copeland, TD, Levine, SS, Lee, JC, Hayes, DN, Shanmugam, KS, Bhattacharjee, A, Biondi, CA, Kay, GF, Hayward, NK, Hess, JL, Meyerson, M (2004) Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. Mol Cell *13*, 587-597.

Hunger, SP, Tkachuk, DC, Amylon, MD, Link, MP, Carroll, AJ, Welborn, JL, Willman, CL, Cleary, ML (1993) HRX involvement in de novo and secondary leukemias with diverse chromosome 11q23 abnormalities. Blood *81*, 3197-3203.

Hurwitz, J, Bresler, A, Diringer, R (1960) The enzymic incorporation of ribonucleotides into polyribonucleotides and the effect of DNA. Biochem Biophys Res Commun *3*, 15-18.

Inostroza, J, Flores, O, Reinberg, D (1991) Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of general transcription factor IIE. J Biol Chem *266*, 9304-9308.

Izban, MG, Luse, DS (1992a) Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. J Biol Chem *267*, 13647-13655.

Izban, MG, Luse, DS (1992b) The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. Genes Dev *6*, 1342-1356.

Jang, MK, Mochizuki, K, Zhou, M, Jeong, HS, Brady, JN, Ozato, K (2005) The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol Cell *19*, 523-534.

Jishage, M, Malik, S, Wagner, U, Uberheide, B, Ishihama, Y, Hu, X, Chait, BT, Gnatt, A, Ren, B, Roeder, RG (2012) Transcriptional regulation by Pol II(G) involving mediator and competitive interactions of Gdown1 and TFIIF with Pol II. Mol Cell *45*, 51-63.

Johnstone, RW, Gerber, M, Landewe, T, Tollefson, A, Wold, WS, Shilatifard, A (2001) Functional analysis of the leukemia protein ELL: evidence for a role in the regulation of cell growth and survival. Mol Cell Biol *21*, 1672-1681.

Jones, KA, Peterlin, BM (1994) Control of RNA initiation and elongation at the HIV-1 promoter. Annu Rev Biochem *63*, 717-743.

Kagey, MH, Newman, JJ, Bilodeau, S, Zhan, Y, Orlando, DA, van Berkum, NL, Ebmeier, CC, Goossens, J, Rahl, PB, Levine, SS, Taatjes, DJ, Dekker, J, Young, RA (2010) Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kelly, WG, Dahmus, ME, Hart, GW (1993) RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. J Biol Chem *268*, 10416-10424.

Kent, WJ, Sugnet, CW, Furey, TS, Roskin, KM, Pringle, TH, Zahler, AM, Haussler, D (2002) The human genome browser at UCSC. Genome Res *12*, 996-1006.

Kettenberger, H, Armache, KJ, Cramer, P (2003) Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. Cell *114*, 347-357.

Khoueiry, P, Rothbacher, U, Ohtsuka, Y, Daian, F, Frangulian, E, Roure, A, Dubchak, I, Lemaire, P (2010) A cis-regulatory signature in ascidians and flies, independent of transcription factor binding sites. Curr Biol *20*, 792-802.

Kim, H, Erickson, B, Luo, W, Seward, D, Graber, JH, Pollock, DD, Megee, PC, Bentley, DL (2010) Gene-specific RNA polymerase II phosphorylation and the CTD code. Nat Struct Mol Biol *17*, 1279-1286.

Kim, JB, Sharp, PA (2001) Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. J Biol Chem *276*, 12317-12323.

Kim, M, Vasiljeva, L, Rando, OJ, Zhelkovsky, A, Moore, C, Buratowski, S (2006) Distinct pathways for snoRNA and mRNA termination. Mol Cell *24*, 723-734.

Kim, WY, Dahmus, ME (1989) The major late promoter of adenovirus-2 is accurately transcribed by RNA polymerases IIO, IIA, and IIB. J Biol Chem *264*, 3169-3176.

Knight, SJ, Flannery, AV, Hirst, MC, Campbell, L, Christodoulou, Z, Phelps, SR, Pointon, J, Middleton-Price, HR, Barnicoat, A, Pembrey, ME, et al. (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell *74*, 127-134.

Komarnitsky, P, Cho, EJ, Buratowski, S (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev *14*, 2452-2460.

Komissarova, N, Kashlev, M (1997) Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. Proc Natl Acad Sci U S A *94*, 1755-1760.

Kong, SE, Banks, CA, Shilatifard, A, Conaway, JW, Conaway, RC (2005) ELL-associated factors 1 and 2 are positive regulators of RNA polymerase II elongation factor ELL. Proc Natl Acad Sci U S A *102*, 10094-10098.

Krivtsov, AV, Armstrong, SA (2007) MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer *7*, 823-833.

Krivtsov, AV, Feng, Z, Lemieux, ME, Faber, J, Vempati, S, Sinha, AU, Xia, X, Jesneck, J, Bracken, AP, Silverman, LB, Kutok, JL, Kung, AL, Armstrong, SA (2008) H3K79 methylation profiles define murine and human MLL-AF4 leukemias. Cancer Cell *14*, 355-368.

Krivtsov, AV, Twomey, D, Feng, Z, Stubbs, MC, Wang, Y, Faber, J, Levine, JE, Wang, J, Hahn, WC, Gilliland, DG, Golub, TR, Armstrong, SA (2006) Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. Nature *442*, 818-822.

Krogan, NJ, Dover, J, Khorrami, S, Greenblatt, JF, Schneider, J, Johnston, M, Shilatifard, A (2002) COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. J Biol Chem *277*, 10753-10755.

Kroon, E, Krosl, J, Thorsteinsdottir, U, Baban, S, Buchberg, AM, Sauvageau, G (1998) Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. EMBO J *17*, 3714-3725.

Krumm, A, Meulia, T, Brunvand, M, Groudine, M (1992) The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region. Genes Dev *6*, 2201-2213.

Krumm, A, Meulia, T, Groudine, M (1993) Common mechanisms for the control of eukaryotic transcriptional elongation. Bioessays *15*, 659-665.

Kuehner, JN, Pearson, EL, Moore, C (2011) Unravelling the means to an end: RNA polymerase II transcription termination. Nat Rev Mol Cell Biol *12*, 283-294.

Kugel, JF, Goodrich, JA (1998) Promoter escape limits the rate of RNA polymerase II transcription and is enhanced by TFIIE, TFIIH, and ATP on negatively supercoiled DNA. Proc Natl Acad Sci U S A *95*, 9232-9237.

Kumar, AR, Hudson, WA, Chen, W, Nishiuchi, R, Yao, Q, Kersey, JH (2004) Hoxa9 influences the phenotype but not the incidence of Mll-AF9 fusion gene leukemia. Blood *103*, 1823-1828.

Kumar, KP, Akoulitchev, S, Reinberg, D (1998) Promoter-proximal stalling results from the inability to recruit transcription factor IIH to the transcription complex and is a regulated event. Proc Natl Acad Sci U S A *95*, 9767-9772.

Kurosawa, H (2007) Methods for inducing embryoid body formation: in vitro differentiation system of embryonic stem cells. J Biosci Bioeng *103*, 389-398.

Kvon, EZ, Stampfel, G, Yanez-Cuna, JO, Dickson, BJ, Stark, A (2012) HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. Genes Dev *26*, 908-913.

Langmead, B, Trapnell, C, Pop, M, Salzberg, SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Laspia, MF, Rice, AP, Mathews, MB (1989) HIV-1 Tat protein increases transcriptional initiation and stabilizes elongation. Cell *59*, 283-292.

Learned, RM, Learned, TK, Haltiner, MM, Tjian, RT (1986) Human rRNA transcription is modulated by the coordinate binding of two factors to an upstream control element. Cell *45*, 847-857.

Lee, H, Kraus, KW, Wolfner, MF, Lis, JT (1992) DNA sequence requirements for generating paused polymerase at the start of hsp70. Genes Dev *6*, 284-295.

Lee, JH, Skalnik, DG (2005) CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. J Biol Chem *280*, 41725-41731.

Lee, JH, Tate, CM, You, JS, Skalnik, DG (2007) Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. J Biol Chem *282*, 13419-13428.

Lee, S, Roeder, RG, Lee, JW (2009) Roles of histone H3-lysine 4 methyltransferase complexes in NR-mediated gene transcription. Prog Mol Biol Transl Sci *87*, 343-382.

Lee, TI, Johnstone, SE, Young, RA (2006) Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc *1*, 729-748.

Levine, M (2011) Paused RNA polymerase II as a developmental checkpoint. Cell *145*, 502-511.

Li, G, Ruan, X, Auerbach, RK, Sandhu, KS, Zheng, M, Wang, P, Poh, HM, Goh, Y, Lim, J, Zhang, J, Sim, HS, Peh, SQ, Mulawadi, FH, Ong, CT, Orlov, YL, Hong, S, Zhang, Z, Landt, S, Raha, D, Euskirchen, G, Wei, CL, Ge, W, Wang, H, Davis, C, Fisher-Aylor, KI, Mortazavi, A, Gerstein, M, Gingeras, T, Wold, B, Sun, Y, Fullwood, MJ, Cheung, E, Liu, E, Sung, WK, Snyder, M, Ruan, Y (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell *148*, 84-98.

Li, Q, Price, JP, Byers, SA, Cheng, D, Peng, J, Price, DH (2005) Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. J Biol Chem *280*, 28819-28826.

Liang, HL, Nien, CY, Liu, HY, Metzstein, MM, Kirov, N, Rushlow, C (2008) The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature *456*, 400-403.

Lieberman-Aiden, E, van Berkum, NL, Williams, L, Imakaev, M, Ragoczy, T, Telling, A, Amit, I, Lajoie, BR, Sabo, PJ, Dorschner, MO, Sandstrom, R, Bernstein, B, Bender, MA, Groudine, M, Gnirke, A, Stamatoyannopoulos, J, Mirny, LA, Lander, ES, Dekker, J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Liedtke, M, Cleary, ML (2009) Therapeutic targeting of MLL. Blood *113*, 6061-6068.

Lin, C, Garrett, AS, De Kumar, B, Smith, ER, Gogol, M, Seidel, C, Krumlauf, R, Shilatifard, A (2011) Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). Genes Dev *25*, 1486-1498.

Lin, C, Garruss, AS, Luo, Z, Guo, F, Shilatifard, A (2013) The RNA Pol II Elongation Factor Ell3 Marks Enhancers in ES Cells and Primes Future Gene Activation. Cell *152*, 144-156.

Lin, C, Smith, ER, Takahashi, H, Lai, KC, Martin-Brown, S, Florens, L, Washburn, MP, Conaway, JW, Conaway, RC, Shilatifard, A (2010) AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. Mol Cell *37*, 429-437.

Lis, J (1998) Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. Cold Spring Harb Symp Quant Biol *63*, 347-356.

Liu, H, Cheng, EH, Hsieh, JJ (2009) MLL fusions: pathways to leukemia. Cancer Biol Ther *8*, 1204-1211.

Liu, J, Kipreos, ET (2000) Evolution of cyclin-dependent kinases (CDKs) and CDK-activating kinases (CAKs): differential conservation of CAKs in yeast and metazoa. Mol Biol Evol *17*, 1061-1074.

Lu, KP, Zhou, XZ (2007) The prolyl isomerase PIN1: a pivotal new twist in phosphorylation signalling and disease. Nat Rev Mol Cell Biol *8*, 904-916.

Lu, Q, Wallrath, LL, Granok, H, Elgin, SC (1993) (CT)n (GA)n repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the Drosophila hsp26 gene. Mol Cell Biol *13*, 2802-2814.

Luo, RT, Lavau, C, Du, C, Simone, F, Polak, PE, Kawamata, S, Thirman, MJ (2001) The elongation domain of ELL is dispensable but its ELL-associated factor 1 interaction domain is essential for MLL-ELL-induced leukemogenesis. Mol Cell Biol *21*, 5678-5687.

Luo, Z, Lin, C, Guest, E, Garrett, AS, Mohaghegh, N, Swanson, S, Marshall, S, Florens, L, Washburn, MP, Shilatifard, A (2012a) The Super Elongation Complex Family of RNA Polymerase II Elongation Factors: Gene Target Specificity and Transcriptional Output. Mol Cell Biol *32*, 2608-2617.

Luo, Z, Lin, C, Shilatifard, A (2012b) The super elongation complex (SEC) family in transcriptional control. Nat Rev Mol Cell Biol *13*, 543-547.

Maldonado, E, Ha, I, Cortes, P, Weis, L, Reinberg, D (1990) Factors involved in specific transcription by mammalian RNA polymerase II: role of transcription factors IIA, IID, and IIB during formation of a transcription-competent complex. Mol Cell Biol *10*, 6335-6347.

Mancebo, HS, Lee, G, Flygare, J, Tomassini, J, Luu, P, Zhu, Y, Peng, J, Blau, C, Hazuda, D, Price, D, Flores, O (1997) P-TEFb kinase is required for HIV Tat transcriptional activation in vivo and in vitro. Genes & development *11*, 2633-2644.

Markstein, M, Levine, M (2002) Decoding cis-regulatory DNAs in the Drosophila genome. Current opinion in genetics & development *12*, 601-606.

Marshall, NF, Peng, J, Xie, Z, Price, DH (1996) Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. J Biol Chem *271*, 27176-27183.

Marshall, NF, Price, DH (1995) Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J Biol Chem *270*, 12335-12338.

Marson, A, Levine, SS, Cole, MF, Frampton, GM, Brambrink, T, Johnstone, S, Guenther, MG, Johnston, WK, Wernig, M, Newman, J, Calabrese, JM, Dennis, LM, Volkert, TL, Gupta, S, Love, J, Hannett, N, Sharp, PA, Bartel, DP, Jaenisch, R, Young, RA (2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.

Martincic, K, Alkan, SA, Cheatle, A, Borghesi, L, Milcarek, C (2009) Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. Nat Immunol *10*, 1102-1109.

Matsui, T, Segall, J, Weil, PA, Roeder, RG (1980) Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J Biol Chem *255*, 11992-11996.

Max, T, Sogaard, M, Svejstrup, JQ (2007) Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. J Biol Chem *282*, 14113-14120.

Mayer, A, Lidschreiber, M, Siebert, M, Leike, K, Soding, J, Cramer, P (2010) Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol *17*, 1272-1278.

Mazo, AM, Huang, DH, Mozer, BA, Dawid, IB (1990) The trithorax gene, a trans-acting regulator of the bithorax complex in Drosophila, encodes a protein with zinc-binding domains. Proc Natl Acad Sci U S A *87*, 2112-2116.

McCracken, S, Fong, N, Yankulov, K, Ballantyne, S, Pan, G, Greenblatt, J, Patterson, SD, Wickens, M, Bentley, DL (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357-361.

McGinnis, W, Krumlauf, R (1992) Homeobox genes and axial patterning. Cell *68*, 283-302.

McLean, CY, Bristor, D, Hiller, M, Clarke, SL, Schaar, BT, Lowe, CB, Wenger, AM, Bejerano, G (2010) GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.

Meyer, C, Kowarz, E, Hofmann, J, Renneville, A, Zuna, J, Trka, J, Ben Abdelali, R, Macintyre, E, De Braekeleer, E, De Braekeleer, M, Delabesse, E, de Oliveira, MP, Cave, H, Clappier, E, van Dongen, JJ, Balgobind, BV, van den Heuvel-Eibrink, MM, Beverloo, HB, Panzer-Grumayer, R, Teigler-Schlegel, A, Harbott, J, Kjeldsen, E, Schnittger, S, Koehl, U, Gruhn, B, Heidenreich, O, Chan, LC, Yip, SF, Krzywinski, M, Eckert, C, Moricke, A, Schrappe, M, Alonso, CN, Schafer, BW, Krauter, J, Lee, DA, Zur Stadt, U, Te Kronnie, G, Sutton, R, Izraeli, S, Trakhtenbrot, L, Lo Nigro, L, Tsaur, G, Fechina, L, Szczepanski, T, Strehl, S, Ilencikova, D, Molkentin, M, Burmeister, T, Dingermann, T, Klingebiel, T, Marschalek, R (2009) New insights to the MLL recombinome of acute leukemias. Leukemia *23*, 1490-1499.

Meyer, N, Penn, LZ (2008) Reflecting on 25 years with MYC. Nat Rev Cancer *8*, 976-990.

Mikkelsen, TS, Ku, M, Jaffe, DB, Issac, B, Lieberman, E, Giannoukos, G, Alvarez, P, Brockman, W, Kim, TK, Koche, RP, Lee, W, Mendenhall, E, O'Donovan, A, Presser, A, Russ, C, Xie, X, Meissner, A, Wernig, M, Jaenisch, R, Nusbaum, C, Lander, ES, Bernstein, BE (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Miller, T, Krogan, NJ, Dover, J, Erdjument-Bromage, H, Tempst, P, Johnston, M, Greenblatt, JF, Shilatifard, A (2001) COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc Natl Acad Sci U S A *98*, 12902-12907.

Miller, T, Williams, K, Johnstone, RW, Shilatifard, A (2000) Identification, cloning, expression, and biochemical characterization of the testis-specific RNA polymerase II elongation factor ELL3. J Biol Chem *275*, 32052-32056.

Milne, TA, Briggs, SD, Brock, HW, Martin, ME, Gibbs, D, Allis, CD, Hess, JL (2002) MLL targets SET domain methyltransferase activity to Hox gene promoters. Mol Cell *10*, 1107-1117.

Milne, TA, Kim, J, Wang, GG, Stadler, SC, Basrur, V, Whitcomb, SJ, Wang, Z, Ruthenburg, AJ, Elenitoba-Johnson, KS, Roeder, RG, Allis, CD (2010) Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. Mol Cell *38*, 853-863.

Milne, TA, Martin, ME, Brock, HW, Slany, RK, Hess, JL (2005) Leukemogenic MLL fusion proteins bind across a broad region of the Hox a9 locus, promoting transcription and multiple histone modifications. Cancer Res *65*, 11367-11374.

Mishiro, T, Ishihara, K, Hino, S, Tsutsumi, S, Aburatani, H, Shirahige, K, Kinoshita, Y, Nakao, M (2009) Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. EMBO J *28*, 1234-1245.

Mohan, M, Herz, HM, Takahashi, YH, Lin, C, Lai, KC, Zhang, Y, Washburn, MP, Florens, L, Shilatifard, A (2010a) Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). Genes Dev *24*, 574-589.

Mohan, M, Lin, C, Guest, E, Shilatifard, A (2010b) Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. Nat Rev Cancer *10*, 721-728.

Montavon, T, Soshnikova, N, Mascrez, B, Joye, E, Thevenet, L, Splinter, E, de Laat, W, Spitz, F, Duboule, D (2011) A regulatory archipelago controls Hox genes transcription in digits. Cell *147*, 1132-1145.

Moorman, C, Sun, LV, Wang, J, de Wit, E, Talhout, W, Ward, LD, Greil, F, Lu, XJ, White, KP, Bussemaker, HJ, van Steensel, B (2006) Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. Proc Natl Acad Sci U S A *103*, 12027-12032.

Morris, DP, Greenleaf, AL (2000) The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem *275*, 39935-39943.

Mueller, D, Bach, C, Zeisig, D, Garcia-Cuellar, MP, Monroe, S, Sreekumar, A, Zhou, R, Nesvizhskii, A, Chinnaiyan, A, Hess, JL, Slany, RK (2007) A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. Blood *110*, 4445-4454.

Mueller, D, Garcia-Cuellar, MP, Bach, C, Buhl, S, Maethner, E, Slany, RK (2009) Misguided transcriptional elongation causes mixed lineage leukemia. PLoS Biol *7*, e1000249.

Muntean, AG, Giannola, D, Udager, AM, Hess, JL (2008) The PHD fingers of MLL block MLL fusion protein-mediated transformation. Blood *112*, 4690-4693.

Muntean, AG, Tan, J, Sitwala, K, Huang, Y, Bronstein, J, Connelly, JA, Basrur, V, Elenitoba-Johnson, KS, Hess, JL (2010) The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis. Cancer Cell *17*, 609-621.

Muse, GW, Gilchrist, DA, Nechaev, S, Shah, R, Parker, JS, Grissom, SF, Zeitlinger, J, Adelman, K (2007) RNA polymerase is poised for activation across the genome. Nat Genet *39*, 1507-1511.

Nag, A, Narsinh, K, Martinson, HG (2007) The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. Nat Struct Mol Biol *14*, 662-669.

Nakamura, T, Mori, T, Tada, S, Krajewski, W, Rozovskaia, T, Wassell, R, Dubois, G, Mazo, A, Croce, CM, Canaani, E (2002) ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. Mol Cell *10*, 1119-1128.

Nasmyth, K, Haering, CH (2009) Cohesin: its roles and mechanisms. Annu Rev Genet *43*, 525-558.

Nativio, R, Wendt, KS, Ito, Y, Huddleston, JE, Uribe-Lewis, S, Woodfine, K, Krueger, C, Reik, W, Peters, JM, Murrell, A (2009) Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. PLoS Genet *5*, e1000739.

Nechaev, S, Adelman, K (2008) Promoter-proximal Pol II: when stalling speeds things up. Cell Cycle *7*, 1539-1544.

Nechaev, S, Adelman, K (2011) Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. Biochimica et biophysica acta *1809*, 34-45.

Nechaev, S, Fargo, DC, dos Santos, G, Liu, L, Gao, Y, Adelman, K (2010) Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science *327*, 335-338.

Neufeld, TP, Tang, AH, Rubin, GM (1998) A genetic screen to identify components of the sina signaling pathway in Drosophila eye development. Genetics *148*, 277-286.

Newman, JJ, Young, RA (2010) Connecting transcriptional control to chromosome structure and human disease. Cold Spring Harb Symp Quant Biol *75*, 227-235.

Nien, CY, Liang, HL, Butcher, S, Sun, Y, Fu, S, Gocha, T, Kirov, N, Manak, JR, Rushlow, C (2011) Temporal coordination of gene networks by Zelda in the early Drosophila embryo. PLoS Genet *7*, e1002339.

Noonan, JP, McCallion, AS (2010) Genomics of long-range regulatory elements. Annu Rev Genomics Hum Genet *11*, 1-23.

Nudler, E, Mustaev, A, Lukhtanov, E, Goldfarb, A (1997) The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. Cell *89*, 33-41.

O'Brien, T, Wilkins, RC, Giardina, C, Lis, JT (1995) Distribution of GAGA protein on Drosophila genes in vivo. Genes Dev *9*, 1098-1110.

Ohkuma, Y, Sumimoto, H, Horikoshi, M, Roeder, RG (1990) Factors involved in specific transcription by mammalian RNA polymerase II: purification and characterization of general transcription factor TFIIE. Proc Natl Acad Sci U S A *87*, 9163-9167.

Okada, Y, Feng, Q, Lin, Y, Jiang, Q, Li, Y, Coffield, VM, Su, L, Xu, G, Zhang, Y (2005) hDOT1L links histone methylation to leukemogenesis. Cell *121*, 167-178.

Orphanides, G, Reinberg, D (2002) A unified theory of gene expression. Cell *108*, 439-451.

Osheim, YN, Miller, OL, Jr., Beyer, AL (1985) RNP particles at splice junction sequences on Drosophila chorion transcripts. Cell *43*, 143-151.

Palancade, B, Marshall, NF, Tremeau-Bravard, A, Bensaude, O, Dahmus, ME, Dubois, MF (2004) Dephosphorylation of RNA polymerase II by CTD-phosphatase FCP1 is inhibited by phospho-CTD associating proteins. J Mol Biol *335*, 415-424.

Palermo, CM, Bennett, CA, Winters, AC, Hemenway, CS (2008) The AF4-mimetic peptide, PFWT, induces necrotic cell death in MV4-11 leukemia cells. Leuk Res *32*, 633-642.

Park, NJ, Tsao, DC, Martinson, HG (2004) The two steps of poly(A)-dependent termination, pausing and release, can be uncoupled by truncation of the RNA polymerase II carboxyl-terminal repeat domain. Mol Cell Biol *24*, 4092-4103.

Parker, CS, Jaehning, JA, Roeder, RG (1978) Faithful gene transcription by eukaryotic RNA polymerases in reconstructed systems. Cold Spring Harb Symp Quant Biol *42 Pt 1*, 577-587.

Pascual-Garcia, P, Govind, CK, Queralt, E, Cuenca-Bono, B, Llopis, A, Chavez, S, Hinnebusch, AG, Rodriguez-Navarro, S (2008) Sus1 is recruited to coding regions and functions during transcription elongation in association with SAGA and TREX2. Genes Dev *22*, 2811-2822.

Peterlin, BM, Price, DH (2006) Controlling the elongation phase of transcription with P-TEFb. Mol Cell *23*, 297-305.

Phatnani, HP, Greenleaf, AL (2006) Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev *20*, 2922-2936.

Phatnani, HP, Jones, JC, Greenleaf, AL (2004) Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome. Biochemistry *43*, 15702-15719.

Ping, YH, Rana, TM (2001) DSIF and NELF interact with RNA polymerase II elongation complex and HIV-1 Tat stimulates P-TEFb-mediated phosphorylation of RNA polymerase II and DSIF during transcription elongation. J Biol Chem *276*, 12951-12958.

Polak, PE, Simone, F, Kaberlein, JJ, Luo, RT, Thirman, MJ (2003) ELL and EAF1 are Cajal body components that are disrupted in MLL-ELL leukemia. Mol Biol Cell *14*, 1517-1528.

Popperl, H, Bienz, M, Studer, M, Chan, SK, Aparicio, S, Brenner, S, Mann, RS, Krumlauf, R (1995) Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx. Cell *81*, 1031-1042.

Price, DH, Sluder, AE, Greenleaf, AL (1989) Dynamic interaction between a Drosophila transcription factor and RNA polymerase II. Mol Cell Biol *9*, 1465-1475.

Pui, CH, Behm, FG, Raimondi, SC, Dodge, RK, George, SL, Rivera, GK, Mirro, J, Jr., Kalwinsky, DK, Dahl, GV, Murphy, SB (1989) Secondary acute myeloid leukemia in children treated for acute lymphoid leukemia. N Engl J Med *321*, 136-142.

Rada-Iglesias, A, Bajpai, R, Swigut, T, Brugmann, SA, Flynn, RA, Wysocka, J (2011) A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Rahl, PB, Lin, CY, Seila, AC, Flynn, RA, McCuine, S, Burge, CB, Sharp, PA, Young, RA (2010) c-Myc regulates transcriptional pause release. Cell *141*, 432-445.

Ranish, JA, Hahn, S, Lu, Y, Yi, EC, Li, XJ, Eng, J, Aebersold, R (2004) Identification of TFB5, a new component of general transcription and DNA repair factor IIH. Nat Genet *36*, 707-713.

Raschke, EE, Albert, T, Eick, D (1999) Transcriptional regulation of the Ig kappa gene by promoter-proximal pausing of RNA polymerase II. J Immunol *163*, 4375-4382.

Rasmussen, EB, Lis, JT (1993) In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc Natl Acad Sci U S A *90*, 7923-7927.

Rasmussen, EB, Lis, JT (1995) Short transcripts of the ternary complex provide insight into RNA polymerase II elongational pausing. J Mol Biol *252*, 522-535.

Reed, R (2003) Coupling transcription, splicing and mRNA export. Curr Opin Cell Biol *15*, 326-331.

Renner, DB, Yamaguchi, Y, Wada, T, Handa, H, Price, DH (2001) A highly purified RNA polymerase II elongation control system. J Biol Chem *276*, 42601-42609.

Rhee, HS, Pugh, BF (2012) Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295-301.

Ribar, B, Prakash, L, Prakash, S (2007) ELA1 and CUL3 are required along with ELC1 for RNA polymerase II polyubiquitylation and degradation in DNA-damaged yeast cells. Mol Cell Biol *27*, 3211-3216.

Ringrose, L, Paro, R (2004) Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu Rev Genet *38*, 413-443.

Roeder, RG (1996) The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci *21*, 327-335.

Roeder, RG (2005) Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS Lett *579*, 909-915.

Roeder, RG, Rutter, WJ (1969) Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature *224*, 234-237.

Roeder, RG, Rutter, WJ (1970a) Multiple ribonucleic acid polymerases and ribonucleic acid synthesis during sea urchin development. Biochemistry *9*, 2543-2553.

Roeder, RG, Rutter, WJ (1970b) Specific nucleolar and nucleoplasmic RNA polymerases. Proc Natl Acad Sci U S A *65*, 675-682.

Rollins, RA, Morcillo, P, Dorsett, D (1999) Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics *152*, 577-593.

Romano, G, Giordano, A (2008) Role of the cyclin-dependent kinase 9-related pathway in mammalian gene expression and human diseases. Cell Cycle *7*, 3664-3668.

Rougvie, AE, Lis, JT (1988) The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell *54*, 795-804.

Rowley, JD (1998) The critical role of chromosome translocations in human leukemias. Annu Rev Genet *32*, 495-519.

Rowley, JD (2001) Chromosome translocations: dangerous liaisons revisited. Nat Rev Cancer *1*, 245-250.

Rowley, JD (2008) Chromosomal translocations: revisited yet again. Blood *112*, 2183-2189.

Roy, R, Adamczewski, JP, Seroz, T, Vermeulen, W, Tassan, JP, Schaeffer, L, Nigg, EA, Hoeijmakers, JH, Egly, JM (1994) The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. Cell *79*, 1093-1101.

Rozovskaia, T, Feinstein, E, Mor, O, Foa, R, Blechman, J, Nakamura, T, Croce, CM, Cimino, G, Canaani, E (2001) Upregulation of Meis1 and HoxA9 in acute lymphocytic leukemias with the t(4 : 11) abnormality. Oncogene *20*, 874-878.

Sakai, Y, Meno, C, Fujii, H, Nishino, J, Shiratori, H, Saijoh, Y, Rossant, J, Hamada, H (2001) The retinoic acid-inactivating enzyme CYP26 is essential for establishing an uneven distribution of retinoic acid along the anterio-posterior axis within the mouse embryo. Genes Dev *15*, 213-225.

Samuels, M, Fire, A, Sharp, PA (1982) Separation and characterization of factors mediating accurate transcription by RNA polymerase II. J Biol Chem *257*, 14419-14427.

Samuels, M, Sharp, PA (1986) Purification and characterization of a specific RNA polymerase II transcription factor. J Biol Chem *261*, 2003-2013.

Sandaltzopoulos, R, Mitchelmore, C, Bonte, E, Wall, G, Becker, PB (1995) Dual regulation of the Drosophila hsp26 promoter in vitro. Nucleic Acids Res *23*, 2479-2487.

Schroeder, SC, Schwer, B, Shuman, S, Bentley, D (2000) Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes Dev *14*, 2435-2440.

Schulze, JM, Jackson, J, Nakanishi, S, Gardner, JM, Hentrich, T, Haug, J, Johnston, M, Jaspersen, SL, Kobor, MS, Shilatifard, A (2009) Linking cell cycle to histone modifications: SBF and H2B monoubiquitination machinery and cell-cycle regulation of H3K79 dimethylation. Mol Cell *35*, 626-641.

Sekimizu, K, Kobayashi, N, Mizuno, D, Natori, S (1976) Purification of a factor from Ehrlich ascites tumor cells specifically stimulating RNA polymerase II. Biochemistry *15*, 5064-5070.

Sentenac, A (1985) Eukaryotic RNA polymerases. CRC Crit Rev Biochem *18*, 31-90.

Serizawa, H, Makela, TP, Conaway, JW, Conaway, RC, Weinberg, RA, Young, RA (1995) Association of Cdk-activating kinase subunits with transcription factor TFIIH. Nature *374*, 280-282.

Shandilya, J, Roberts, SG (2012) The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. Biochim Biophys Acta *1819*, 391-400.

Shearn, A (1989) The ash-1, ash-2 and trithorax genes of Drosophila melanogaster are functionally related. Genetics *121*, 517-525.

Shell, SA, Martincic, K, Tran, J, Milcarek, C (2007) Increased phosphorylation of the carboxyl-terminal domain of RNA polymerase II and loading of polyadenylation and cotranscriptional factors contribute to regulation of the ig heavy chain mRNA in plasma cells. J Immunol *179*, 7663-7673.

Shiekhattar, R, Mermelstein, F, Fisher, RP, Drapkin, R, Dynlacht, B, Wessling, HC, Morgan, DO, Reinberg, D (1995) Cdk-activating kinase complex is a component of human transcription factor TFIIH. Nature *374*, 283-287.

Shilatifard, A (1998) Factors regulating the transcriptional elongation activity of RNA polymerase II. FASEB J *12*, 1437-1446.

Shilatifard, A (2012) The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. Annu Rev Biochem *81*, 65-95.

Shilatifard, A, Conaway, RC, Conaway, JW (2003) The RNA polymerase II elongation complex. Annu Rev Biochem *72*, 693-715.

Shilatifard, A, Lane, WS, Jackson, KW, Conaway, RC, Conaway, JW (1996) An RNA polymerase II elongation factor encoded by the human ELL gene. Science *271*, 1873-1876.

Shinobu, N, Maeda, T, Aso, T, Ito, T, Kondo, T, Koike, K, Hatakeyama, M (1999) Physical interaction and functional antagonism between the RNA polymerase II elongation factor ELL and p53. J Biol Chem *274*, 17003-17010.

Showe, LC, Croce, CM (1987) The role of chromosomal translocations in B- and T-cell neoplasia. Annu Rev Immunol *5*, 253-277.

Shuman, S (1997) Origins of mRNA identity: capping enzymes bind to the phosphorylated C-terminal domain of RNA polymerase II. Proc Natl Acad Sci U S A *94*, 12758-12760.

Sikorski, TW, Buratowski, S (2009) The basal initiation machinery: beyond the general transcription factors. Curr Opin Cell Biol *21*, 344-351.

Simone, F, Luo, RT, Polak, PE, Kaberlein, JJ, Thirman, MJ (2003) ELL-associated factor 2 (EAF2), a functional homolog of EAF1 with alternative ELL binding properties. Blood *101*, 2355-2362.

Simone, F, Polak, PE, Kaberlein, JJ, Luo, RT, Levitan, DA, Thirman, MJ (2001) EAF1, a novel ELL-associated factor that is delocalized by expression of the MLL-ELL fusion protein. Blood *98*, 201-209.

Sims, RJ, 3rd, Belotserkovskaya, R, Reinberg, D (2004) Elongation by RNA polymerase II: the short and long of it. Genes Dev *18*, 2437-2468.

Sims, RJ, 3rd, Rojas, LA, Beck, D, Bonasio, R, Schuller, R, Drury, WJ, 3rd, Eick, D, Reinberg, D (2011) The C-terminal domain of RNA polymerase II is modified by site-specific methylation. Science *332*, 99-103.

Skibbens, RV (2009) Establishment of sister chromatid cohesion. Curr Biol *19*, R1126-1132.

Slany, RK, Lavau, C, Cleary, ML (1998) The oncogenic capacity of HRX-ENL requires the transcriptional transactivation activity of ENL and the DNA binding motifs of HRX. Mol Cell Biol *18*, 122-129.

Smale, ST, Kadonaga, JT (2003) The RNA polymerase II core promoter. Annu Rev Biochem *72*, 449-479.

Smale, ST, Tjian, R (1985) Transcription of herpes simplex virus tk sequences under the control of wild-type and mutant human RNA polymerase I promoters. Mol Cell Biol *5*, 352-362.

Smith, E, Lin, C, Shilatifard, A (2011a) The super elongation complex (SEC) and MLL in development and disease. Genes Dev *25*, 661-672.

Smith, E, Shilatifard, A (2010) The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. Mol Cell *40*, 689-701.

Smith, ER, Lin, C, Garrett, AS, Thornton, J, Mohaghegh, N, Hu, D, Jackson, J, Saraf, A, Swanson, SK, Seidel, C, Florens, L, Washburn, MP, Eissenberg, JC, Shilatifard, A (2011b) The little elongation complex regulates small nuclear RNA transcription. Mol Cell *44*, 954-965.

Smith, ER, Winter, B, Eissenberg, JC, Shilatifard, A (2008) Regulation of the transcriptional activity of poised RNA polymerase II by the elongation factor ELL. Proc Natl Acad Sci U S A *105*, 8575-8579.

Smyth, GK, Michaud, J, Scott, HS (2005) Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics *21*, 2067-2075.

So, CW, Cleary, ML (2004) Dimerization: a versatile switch for oncogenesis. Blood *104*, 919-922.

So, CW, Karsunky, H, Passegue, E, Cozzio, A, Weissman, IL, Cleary, ML (2003a) MLL-GAS7 transforms multipotent hematopoietic progenitors and induces mixed lineage leukemias in mice. Cancer Cell *3*, 161-171.

So, CW, Lin, M, Ayton, PM, Chen, EH, Cleary, ML (2003b) Dimerization contributes to oncogenic activation of MLL chimeras in acute leukemias. Cancer Cell *4*, 99-110.

Sobhian, B, Laguette, N, Yatim, A, Nakamura, M, Levy, Y, Kiernan, R, Benkirane, M (2010) HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. Mol Cell *38*, 439-451.

Somervaille, TC, Cleary, ML (2006) Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia. Cancer Cell *10*, 257-268.

Somervaille, TC, Matheny, CJ, Spencer, GJ, Iwasaki, M, Rinn, JL, Witten, DM, Chang, HY, Shurtleff, SA, Downing, JR, Cleary, ML (2009) Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. Cell Stem Cell *4*, 129-140.

Spencer, CA, Groudine, M (1990) Molecular analysis of the c-myc transcription elongation block. Implications for the generation of Burkitt's lymphoma. Ann N Y Acad Sci *599*, 12-28.

Srinivasan, RS, Nesbit, JB, Marrero, L, Erfurth, F, LaRussa, VF, Hemenway, CS (2004) The synthetic peptide PFWT disrupts AF4-AF9 protein complexes and induces apoptosis in t(4;11) leukemia cells. Leukemia *18*, 1364-1372.

Steitz, TA (1998) A mechanism for all polymerases. Nature *391*, 231-232.

Stevens, A (1960) Incorporation of the adenine ribonucleotide into RNA by cell fractions from E. coli B. Biochem Biophys Res Commun *3*, 92-96.

Steward, MM, Lee, JS, O'Donovan, A, Wyatt, M, Bernstein, BE, Shilatifard, A (2006) Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. Nat Struct Mol Biol *13*, 852-854.

Stock, JK, Giadrossi, S, Casanova, M, Brookes, E, Vidal, M, Koseki, H, Brockdorff, N, Fisher, AG, Pombo, A (2007) Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat Cell Biol *9*, 1428-1435.

Studer, M, Gavalas, A, Marshall, H, Ariza-McNaughton, L, Rijli, FM, Chambon, P, Krumlauf, R (1998) Genetic interactions between Hoxa1 and Hoxb1 reveal new roles in regulation of early hindbrain patterning. Development *125*, 1025-1036.

Su, MA, Wisotzkey, RG, Newfeld, SJ (2001) A screen for modifiers of decapentaplegic mutant phenotypes identifies lilliputian, the only member of the Fragile-X/Burkitt's Lymphoma family of transcription factors in Drosophila melanogaster. Genetics *157*, 717-725.

Sun, X, Zhang, Y, Cho, H, Rickert, P, Lees, E, Lane, W, Reinberg, D (1998) NAT, a human complex containing Srb polypeptides that functions as a negative regulator of activated transcription. Mol Cell *2*, 213-222.

Super, HJ, McCabe, NR, Thirman, MJ, Larson, RA, Le Beau, MM, Pedersen-Bjergaard, J, Philip, P, Diaz, MO, Rowley, JD (1993) Rearrangements of the MLL gene in therapy-related acute myeloid leukemia in patients previously treated with agents targeting DNA-topoisomerase II. Blood *82*, 3705-3711.

Suraweera, A, Lim, Y, Woods, R, Birrell, GW, Nasim, T, Becherel, OJ, Lavin, MF (2009) Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. Hum Mol Genet *18*, 3384-3396.

Takahashi, H, Parmely, TJ, Sato, S, Tomomori-Sato, C, Banks, CA, Kong, SE, Szutorisz, H, Swanson, SK, Martin-Brown, S, Washburn, MP, Florens, L, Seidel, CW, Lin, C, Smith, ER, Shilatifard, A, Conaway, RC, Conaway, JW (2011) Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell *146*, 92-104.

Takeda, S, Chen, DY, Westergard, TD, Fisher, JK, Rubens, JA, Sasagawa, S, Kan, JT, Korsmeyer, SJ, Cheng, EH, Hsieh, JJ (2006) Proteolysis of MLL family proteins is essential for taspase1-orchestrated cell cycle progression. Genes Dev *20*, 2397-2409.

Tan, S, Aso, T, Conaway, RC, Conaway, JW (1994) Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II. J Biol Chem *269*, 25684-25691.

Tanese, N, Pugh, BF, Tjian, R (1991) Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. Genes Dev *5*, 2212-2224.

Tang, AH, Neufeld, TP, Rubin, GM, Muller, HA (2001) Transcriptional regulation of cytoskeletal functions and segmentation by a novel maternal pair-rule gene, lilliputian. Development *128*, 801-813.

Tang, XH, Gudas, LJ (2011) Retinoids, retinoic acid receptors, and cancer. Annu Rev Pathol *6*, 345-364.

Taub, R, Kirsch, I, Morton, C, Lenoir, G, Swan, D, Tronick, S, Aaronson, S, Leder, P (1982) Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. Proc Natl Acad Sci U S A *79*, 7837-7841.

Tenney, K, Gerber, M, Ilvarsonn, A, Schneider, J, Gause, M, Dorsett, D, Eissenberg, JC, Shilatifard, A (2006) Drosophila Rtf1 functions in histone methylation, gene expression, and Notch signaling. Proc Natl Acad Sci U S A *103*, 11970-11974.

Terranova, R, Agherbi, H, Boned, A, Meresse, S, Djabali, M (2006) Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. Proc Natl Acad Sci U S A *103*, 6629-6634.

Thiebaut, M, Kisseleva-Romanova, E, Rougemaille, M, Boulay, J, Libri, D (2006) Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. Mol Cell *23*, 853-864.

Thiel, AT, Blessington, P, Zou, T, Feather, D, Wu, X, Yan, J, Zhang, H, Liu, Z, Ernst, P, Koretzky, GA, Hua, X (2010) MLL-AF9-induced leukemogenesis requires coexpression of the wild-type Mll allele. Cancer Cell *17*, 148-159.

Thirman, MJ, Levitan, DA, Kobayashi, H, Simon, MC, Rowley, JD (1994) Cloning of ELL, a gene that fuses to MLL in a t(11;19)(q23;p13.1) in acute myeloid leukemia. Proc Natl Acad Sci U S A *91*, 12110-12114.

Thomas, MC, Chiang, CM (2006) The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol *41*, 105-178.

Tietjen, JR, Zhang, DW, Rodriguez-Molina, JB, White, BE, Akhtar, MS, Heidemann, M, Li, X, Chapman, RD, Shokat, K, Keles, S, Eick, D, Ansari, AZ (2010) Chemical-genomic dissection of the CTD code. Nat Struct Mol Biol *17*, 1154-1161.

Tkachuk, DC, Kohler, S, Cleary, ML (1992) Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias. Cell *71*, 691-700.

Trapnell, C, Pachter, L, Salzberg, SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trapnell, C, Williams, BA, Pertea, G, Mortazavi, A, Kwan, G, van Baren, MJ, Salzberg, SL, Wold, BJ, Pachter, L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

Tsukiyama, T, Wu, C (1995) Purification and properties of an ATP-dependent nucleosome remodeling factor. Cell *83*, 1011-1020.

von Bergh, AR, Beverloo, HB, Rombout, P, van Wering, ER, van Weel, MH, Beverstock, GC, Kluin, PM, Slater, RM, Schuuring, E (2002) LAF4, an AF4-related gene, is fused to MLL in infant acute lymphoblastic leukemia. Genes Chromosomes Cancer *35*, 92-96.

Wada, T, Takagi, T, Yamaguchi, Y, Ferdous, A, Imai, T, Hirose, S, Sugimoto, S, Yano, K, Hartzog, GA, Winston, F, Buratowski, S, Handa, H (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.

Wall, G, Varga-Weisz, PD, Sandaltzopoulos, R, Becker, PB (1995) Chromatin remodeling by GAGA factor and heat shock factor at the hypersensitive Drosophila hsp26 promoter in vitro. EMBO J *14*, 1727-1736.

Wang, P, Lin, C, Smith, ER, Guo, H, Sanderson, BW, Wu, M, Gogol, M, Alexander, T, Seidel, C, Wiedemann, LM, Ge, K, Krumlauf, R, Shilatifard, A (2009) Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. Mol Cell Biol *29*, 6074-6085.

Wang, Y, Krivtsov, AV, Sinha, AU, North, TE, Goessling, W, Feng, Z, Zon, LI, Armstrong, SA (2010a) The Wnt/beta-catenin pathway is required for the development of leukemia stem cells in AML. Science *327*, 1650-1653.

Wang, Z, Song, J, Milne, TA, Wang, GG, Li, H, Allis, CD, Patel, DJ (2010b) Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression. Cell *141*, 1183-1194.

Wedeen, C, Harding, K, Levine, M (1986) Spatial regulation of Antennapedia and bithorax gene expression by the Polycomb locus in Drosophila. Cell *44*, 739-748.

Wei, P, Garber, ME, Fang, SM, Fischer, WH, Jones, KA (1998) A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. Cell *92*, 451-462.

Weil, PA, Luse, DS, Segall, J, Roeder, RG (1979) Selective and accurate initiation of transcription at the Ad2 major late promotor in a soluble system dependent on purified RNA polymerase II and DNA. Cell *18*, 469-484.

Weiss, SB, Gladstone, l (1959) A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. J Am Chem Soc *81*, 4118-4119.

Weiss, SB, Nakamoto, T (1961) Net synthesis of ribonucleic acid with a microbial enzyme requiring deoxyribonucleic acid and four ribonucleoside triphosphates. J Biol Chem *236*, PC18-20.

Wen, Y, Shatkin, AJ (1999) Transcription elongation factor hSPT5 stimulates mRNA capping. Genes Dev *13*, 1774-1779.

Wendt, KS, Peters, JM (2009) How cohesin and CTCF cooperate in regulating gene expression. Chromosome Res *17*, 201-214.

Westover, KD, Bushnell, DA, Kornberg, RD (2004) Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. Cell *119*, 481-489.

Wiederschain, D, Kawai, H, Gu, J, Shilatifard, A, Yuan, ZM (2003) Molecular basis of p53 functional inactivation by the leukemic protein MLL-ELL. Mol Cell Biol *23*, 4230-4246.

Wittwer, F, van der Straten, A, Keleman, K, Dickson, BJ, Hafen, E (2001) Lilliputian: an AF4/FMR2-related protein that controls cell identity and cell growth. Development *128*, 791-800.

Workman, JL, Kingston, RE (1998) Alteration of nucleosome structure as a mechanism of transcriptional regulation. Annu Rev Biochem *67*, 545-579.

Woychik, NA, Young, RA (1990) RNA polymerase II: subunit structure and function. Trends Biochem Sci *15*, 347-351.

Wu, CH, Lee, C, Fan, R, Smith, MJ, Yamaguchi, Y, Handa, H, Gilmour, DS (2005) Molecular characterization of Drosophila NELF. Nucleic Acids Res *33*, 1269-1279.

Wu, CH, Yamaguchi, Y, Benjamin, LR, Horvat-Gordon, M, Washinsky, J, Enerly, E, Larsson, J, Lambertsson, A, Handa, H, Gilmour, D (2003) NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. Genes Dev *17*, 1402-1414.

Wu, M, Wang, PF, Lee, JS, Martin-Brown, S, Florens, L, Washburn, M, Shilatifard, A (2008) Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Mol Cell Biol *28*, 7337-7344.

Wu, X, Wilcox, CB, Devasahayam, G, Hackett, RL, Arevalo-Rodriguez, M, Cardenas, ME, Heitman, J, Hanes, SD (2000) The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery. EMBO J *19*, 3727-3738.

Xu, YX, Hirose, Y, Zhou, XZ, Lu, KP, Manley, JL (2003) Pin1 modulates the structure and function of human RNA polymerase II. Genes Dev *17*, 2765-2776.

Yagi, H, Deguchi, K, Aono, A, Tani, Y, Kishimoto, T, Komori, T (1998) Growth disturbance in fetal liver hematopoiesis of Mll-mutant mice. Blood *92*, 108-117.

Yamada, T, Yamaguchi, Y, Inukai, N, Okamoto, S, Mura, T, Handa, H (2006) P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol Cell *21*, 227-237.

Yamaguchi, Y, Inukai, N, Narita, T, Wada, T, Handa, H (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*, 2918-2927.

Yamaguchi, Y, Takagi, T, Wada, T, Yano, K, Furuya, A, Sugimoto, S, Hasegawa, J, Handa, H (1999) NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell *97*, 41-51.

Yan, Q, Moreland, RJ, Conaway, JW, Conaway, RC (1999) Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. J Biol Chem *274*, 35668-35675.

Yang, Z, He, N, Zhou, Q (2008) Brd4 recruits P-TEFb to chromosomes at late mitosis to promote G1 gene expression and cell cycle progression. Mol Cell Biol *28*, 967-976.

Yang, Z, Yik, JH, Chen, R, He, N, Jang, MK, Ozato, K, Zhou, Q (2005) Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Mol Cell *19*, 535-545.

Yasukawa, T, Kamura, T, Kitajima, S, Conaway, RC, Conaway, JW, Aso, T (2008) Mammalian Elongin A complex mediates DNA-damage-induced ubiquitylation and degradation of Rpb1. EMBO J *27*, 3256-3266.

Yeung, J, Esposito, MT, Gandillet, A, Zeisig, BB, Griessinger, E, Bonnet, D, So, CW (2010) beta-Catenin mediates the establishment and drug resistance of MLL leukemic stem cells. Cancer Cell *18*, 606-618.

Yokoyama, A, Cleary, ML (2008) Menin critically links MLL proteins with LEDGF on cancer-associated target genes. Cancer Cell *14*, 36-46.

Yokoyama, A, Ficara, F, Murphy, MJ, Meisel, C, Naresh, A, Kitabayashi, I, Cleary, ML (2011) Proteolytically cleaved MLL subunits are susceptible to distinct degradation pathways. J Cell Sci *124*, 2208-2219.

Yokoyama, A, Kitabayashi, I, Ayton, PM, Cleary, ML, Ohki, M (2002) Leukemia proto-oncoprotein MLL is proteolytically processed into 2 fragments with opposite transcriptional properties. Blood *100*, 3710-3718.

Yokoyama, A, Lin, M, Naresh, A, Kitabayashi, I, Cleary, ML (2010) A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. Cancer Cell *17*, 198-212.

Young, RA (1991) RNA polymerase II. Annu Rev Biochem *60*, 689-715.

Yu, BD, Hess, JL, Horning, SE, Brown, GA, Korsmeyer, SJ (1995) Altered Hox expression and segmental identity in Mll-mutant mice. Nature *378*, 505-508.

Yura, T, Ishihama, A (1979) Genetics of bacterial RNA polymerases. Annu Rev Genet *13*, 59-97.

Zawel, L, Reinberg, D (1993) Initiation of transcription by RNA polymerase II: a multi-step process. Prog Nucleic Acid Res Mol Biol *44*, 67-108.

Zeisig, BB, Milne, T, Garcia-Cuellar, MP, Schreiner, S, Martin, ME, Fuchs, U, Borkhardt, A, Chanda, SK, Walker, J, Soden, R, Hess, JL, Slany, RK (2004) Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular immortalization. Mol Cell Biol *24*, 617-628.

Zeitlinger, J, Stark, A, Kellis, M, Hong, JW, Nechaev, S, Adelman, K, Levine, M, Young, RA (2007) RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet *39*, 1512-1516.

Zeleznik-Le, NJ, Harden, AM, Rowley, JD (1994) 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. Proc Natl Acad Sci U S A *91*, 10610-10614.

Zhang, JA, Mortazavi, A, Williams, BA, Wold, BJ, Rothenberg, EV (2012) Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. Cell *149*, 467-482.

Zhang, W, Xia, X, Reisenauer, MR, Hemenway, CS, Kone, BC (2006) Dot1a-AF9 complex mediates histone H3 Lys-79 hypermethylation and repression of ENaCalpha in an aldosterone-sensitive manner. J Biol Chem *281*, 18059-18068.

Zhang, Y, Rowley, JD (2006) Chromatin structural elements and chromosomal translocations in leukemia. DNA Repair (Amst) *5*, 1282-1297.

Zhou, Q, Li, T, Price, DH (2012) RNA Polymerase II Elongation Control. Annu Rev Biochem.

Zhou, Q, Yik, JH (2006) The Yin and Yang of P-TEFb regulation: implications for human immunodeficiency virus gene expression and global control of cell growth and differentiation. Microbiol Mol Biol Rev *70*, 646-659.

Zhou, QA, Schmidt, MC, Berk, AJ (1991) Requirement for acidic amino acid residues immediately N-terminal to the conserved domain of Saccharomyces cerevisiae TFIID. EMBO J *10*, 1843-1852.

Zhu, Y, Pe'ery, T, Peng, J, Ramanathan, Y, Marshall, N, Marshall, T, Amendt, B, Mathews, MB, Price, DH (1997) Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. Genes & development *11*, 2622-2632.

Zippo, A, Serafini, R, Rocchigiani, M, Pennacchini, S, Krepelova, A, Oliviero, S (2009) Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. Cell *138*, 1122-1136.

Zuber, J, Shi, J, Wang, E, Rappaport, AR, Herrmann, H, Sison, EA, Magoon, D, Qi, J, Blatt, K, Wunderlich, M, Taylor, MJ, Johns, C, Chicas, A, Mulloy, JC, Kogan, SC, Brown, P, Valent, P, Bradner, JE, Lowe, SW, Vakoc, CR (2011) RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature *478*, 524-528.

2556