



The impact of DNA sequence and chromatin on transcription in
Trypanosoma brucei

Der Einfluss der DNA-Sequenz und der Chromatinstruktur auf die
Transkription in *Trypanosoma brucei*

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Infection and Immunity

Submitted by

Carolin Wedel

from

Suhl, Germany

Würzburg, 2018



Submitted on:

Members of the Doctoral Thesis Committee:

Chairperson: Prof. Dr. Thomas Dandekar

Primary Supervisor: Prof. T. Nicolai Siegel, PhD

Supervisor (Second): Prof. Dr. Christian Janzen

Supervisor (Third): Prof. Dr. Klaus Brehm

Date of Public Defense:

Date of Receipt of Certificate:

*Almost all aspects of life are engineered at the molecular level,
and without understanding molecules we can only have a very
sketchy understanding of life itself.*

Francis Crick

Affidavit

I hereby confirm that my thesis entitled ‘The impact of DNA sequence and chromatin on transcription in *Trypanosoma brucei*’ is the result of my own work. I did not receive any help or support from commercial consultants. All sources and/or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Place, Date

Signature

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation „Der Einfluss der DNA-Sequenz und der Chromatinstruktur auf die Transkription in *Trypanosoma brucei*“ eigenständig, d.h. insbesondere selbstständig und ohne Hilfsmittel eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Ort, Datum

Unterschrift

Acknowledgements

Foremost, I would like to express my sincere gratitude to my primary supervisor, Nicolai Siegel, for his guidance, continuous support and patience. Thank you for your constant interest in my research and work, your trust, confidence and encouragement when I needed it the most.

I am also very grateful to Christian Janzen, who kindly agreed to be the second reviewer of this thesis. Thank you for the great input you provided in our meetings and your help during my last months in Würzburg.

I would like to thank my third thesis committee member, Klaus Brehm, for taking the time and valuable discussions during our meetings.

I am very grateful to Konrad Förstner, who not only provided computational analysis for the completion of this thesis, but also through his teaching inspired my flourishing interest in bioinformatics. Thank you for taking the time to always discuss and answer my spontaneous questions.

A very special gratitude goes to the former and present lab members of the Siegel lab. Thank you for the great and warm atmosphere, solidarity, the fun in- and outside of the lab and your friendship. Ramona, thank you for your private as well as professional support and your skillful help with the experiments for this thesis. Ruli, thank you for your computational help and your warm personality. Laura and Ines, thank you for reading a part of the introduction and Laura, thank you for encouraging me to look at things from a different perspective. My deepest gratitude goes to Amelie. I actually cannot thank you enough. Thank you for being who you are and for just simply everything.

I also thank the former and present members of the Janzen lab, Zdenka, Helena, Nicole and Tim for many valuable discussions in our lab meetings. I would also like to thank the whole Tryps-community at Hubland. Thank you for your ideas you provided during our TrypClub meetings and for being my worthy substitute work group in Budweis.

A very special thank you goes to my friends at the institute: Caroline, Charlotte, Clivia, Elise, Emmanuel, Falk, Franzi, Gianluca, Jens, Lars, Malvika, Nina, Usha, Yanjie and Youseff. Your company made time so great and you enriched my life so much. Thank you for the fun, the awesome Tuesday seminars and the very informative lunches. I would also like to thank my mentoring peer group members Grit, Jenny and Julia, for their company and support during the last years.

I would like to thank all the fantastic people at the IMIB for the awesome scientific environment, their helpfulness and solidarity.

My heartfelt appreciation goes to my boyfriend, Tobi, for his interest in my work, his constant support and never-ending patience. Thank you for your encouragement especially during the last stage of the development of this thesis and all your love.

Abschließend gilt mein tiefer Dank meinen Eltern, Großeltern und meinem Bruder. Danke für eure Unterstützung in den vergangenen Jahren. Ohne euch wäre das alles nicht möglich gewesen.

Summary

For cellular viability, transcription is a fundamental process. Hereby, the DNA plays the most elemental and highly versatile role. It has long been known that promoters contain conserved and often well-defined motifs, which dictate the site of transcription initiation by providing binding sites for regulatory proteins. However, research within the last decade revealed that it is promoters lacking conserved promoter motifs and transcribing constitutively expressed genes that constitute the majority of promoters in eukaryotes. While the process of transcription initiation is well studied, whether defined DNA sequence motifs are required for the transcription of constitutively expressed genes in eukaryotes remains unknown.

In the highly divergent protozoan parasite *Trypanosoma brucei*, most of the protein-coding genes are organized in large polycistronic transcription units. The genes within one polycistronic transcription unit are generally unrelated and transcribed by a common transcription start site for which no RNA polymerase II promoter motifs have been identified so far. Thus, it is assumed that transcription initiation is not regulated but how transcription is initiated in *T. brucei* is not known. This study aimed to investigate the requirement of DNA sequence motifs and chromatin structures for transcription initiation in an organism lacking transcriptional regulation.

To this end, I performed a systematic analysis to investigate the dependence of transcription initiation on the DNA sequence. I was able to identify GT-rich promoter elements required for directional transcription initiation and targeted deposition of the histone variant H2A.Z, a conserved component during transcription initiation. Furthermore, nucleosome positioning data in this work provide evidence that sites of transcription initiation are rather characterized by broad regions of open and more accessible chromatin than narrow nucleosome depleted regions as it is the case in other eukaryotes. These findings highlight the importance of chromatin during transcription initiation.

Polycistronic RNA in *T. brucei* is separated by adding an independently transcribed miniexon during *trans*-splicing. The data in this work suggest that nucleosome occupancy plays an important role during RNA maturation by slowing down the progressing polymerase and thereby facilitating the choice of the proper splice site during *trans*-splicing.

Overall, this work investigated the role of the DNA sequence during transcription initiation and nucleosome positioning in a highly divergent eukaryote. Furthermore, the findings shed light on the conservation of the requirement of DNA motifs during transcription initiation and the regulatory potential of chromatin during RNA maturation. The findings improve the understanding of gene expression regulation in *T. brucei*, a eukaryotic parasite lacking transcriptional regulation.

Zusammenfassung

Die Transkription ist ein entscheidender Prozess in der Zelle und die DNA-Sequenz nimmt hierbei eine elementare Rolle ein. Promotoren beinhalten spezifische und konservierte DNA-Sequenzen und vermitteln den Start der Transkription durch die Rekrutierung spezifischer Proteine. Jedoch haben Forschungen im vergangenen Jahrzehnt gezeigt, dass die Mehrzahl der Promotoren in eukaryotischen Genomen keine konservierten Promotormotive aufweisen und häufig konstitutiv exprimierte Gene transkribieren. Obgleich der Prozess der Transkriptionsinitiation im Allgemeinen gut erforscht ist, konnte bisher nicht nachgewiesen werden, ob ein definiertes DNA-Motiv während der Transkription von konstitutiv exprimierten Genes erforderlich ist.

In dem eukaryotischen und einzelligen Parasiten *Trypanosoma brucei* ist die Mehrzahl der proteinkodierenden Gene in lange polycistronische Transkriptionseinheiten arrangiert. Diese werden von einem gemeinsamen Transkriptionsstart durch die RNA Polymerase II transkribiert, allerdings konnten hier bisher keine Promotormotive identifiziert werden. Aus diesem Grund besteht die Annahme, dass Transkription keiner Regulation unterliegt. Allgemein ist der Prozess der Transkriptionsinitiation in *T. brucei* bisher nur wenig verstanden.

Um den Zusammenhang zwischen DNA-Motiven und konstitutiver Genexpression näher zu untersuchen und Schlussfolgerungen über die DNA-Sequenz-Abhängigkeit der Transkriptionsinitiation zu ziehen, habe ich eine systematische Analyse in *T. brucei* durchgeführt. Ich konnte GT-reiche Promotorelemente innerhalb dieser Regionen identifizieren, die sowohl eine gerichtete Transkriptionsinitiation, als auch den gezielten Einbau der Histonvariante H2A.Z in Nukleosomen nahe der Transkriptionsstartstelle vermittelt haben. Des Weiteren zeigten Nukleosomenpositionierungsdaten, dass in Trypanosomen die Transkriptionsstartstellen nicht die charakteristische, nukleosomendepletierte Region, wie für andere Organismen beschrieben, sondern eine offene Chromatinstruktur enthalten. Zusätzlich konnte ich zeigen, dass die Chromatinstruktur eine wichtige Rolle während der mRNA-Prozessierung spielt. In *T. brucei* wird die polycistronische pre-mRNA durch das Anfügen eines Miniexons während des sogenannten *trans*-Splicens in individuelle mRNAs aufgetrennt. Die Daten dieser Arbeit belegen, dass die Anreicherung von Nukleosomen eine Verlangsamung der transkribierenden Polymerase bewirken und sie somit die richtige Wahl der Splicestelle gewährleisten.

Zusammenfassend wurde in dieser Arbeit die Rolle der DNA Sequenz während der Transkriptionsinitiation und Nukleosomenpositionierung in einem divergenten Eukaryoten untersucht. Die Erkenntnisse bringen mehr Licht in die Konservierung der Notwendigkeit eines DNA-Motivs während der Transkriptionsinitiation und das regulatorische Potential der

Chromatinstruktur während der RNA-Reifung. Zudem verbessern sie das Verständnis der Genexpressionsregulation in *T. brucei*, einem eukaryotischen Parasiten, der ohne transkriptionelle Regulation überlebt.

Abbreviation index

Δ	deletion	H2A.Z	variant of H2A
3P-RNA	triphosphate RNA	H2B	histone H2B
A (nucleobase)	adenine	H2B.V	variant of H2B
A (amino acid)	alanine	H3	histone H3
aa	amino acid	H3K4me3	H3 tri-methylated at lysine 4
ac	acetylation	H3K76me	H3 methylated at lysine 76
AF	auxiliary factor	H3K79	H3 lysine 79
Amp	ampicillin	H3.V	variant of H3
ATP	adenosine triphosphate	H4	histone H4
bp	base pair(s)	H4K10ac	H4 acetylated at lysine 10
BDF	bromodomain factor	H4K12	H4 lysine 12
BF	bloodstream form	H4.V	variant of H4
Blas	blastocidin	HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
BLE	phleomycin resistance gene	HF	high fidelity
Bre1	E3 ubiquitin-protein ligase	HPLC	high performance liquid chromatography
BRE	TFIIB recognition element	HAS	helicase-SANT
BREd	downstream BRE	HSVTK	<i>Herpes simplex virus</i> thymidine kinase
BREu	upstream BRE	HYG	hygromycin phosphotransferase
Brm	Brahma	Hygro	hygromycin
BSA	bovine serum albumin	I	isoleucine
BSD	blastocidin S deaminase	ID	intra dermal
Bur1/2	bypass UAS requirement	IFA	Incomplete Freund's adjuvant
C (nucleobase)	cytosine	INO80	inositol requiring 80
C (amino acid)	cysteine	Inr	Initiator
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>	ISWI	imitation switch
cDNA	complementary DNA	K	lysine
CDS	coding sequence	kb	kilo base pairs
CFA	Complete Freund's adjuvant	kDa	kilo Dalton
CHD	chromodomain, helicase, DNA binding	KLH	Keyhole limpet haemocyanin
ChIP	chromatin immunoprecipitation	L	leucine
chr	chromosome	LB	lysogeny broth
COMPASS	Complex Proteins Associated with Set1	M	methionine
CPB	counts per billion	me	methylation
Cre	causes recombination	MITat1.2	Moltano Institute Trypanozoon Antigen Type 1.2
Ct	threshold cycle	MNase	micrococcal nuclease
CTD	C-terminal domain	mRNA	messenger-RNA
D	aspartic acid	MTE	motif ten element
<i>D. melanogaster</i>	<i>Drosophila melanogaster</i>	N	asparagine
DCE	downstream core element	Nap1	Nucleosome assembly protein
DHR	downstream homology region	NCBI	National Center for Biotechnology Information
DNA	deoxyribonucleic acid	NDR	nucleosome depleted region
dNTP	deoxyribonucleoside triphosphate	ndTSR	non-divergent TSR
DOT1	disruptor of telomeric silencing 1	NELF	negative elongation factor
DPE	downstream promoter element	NEO	aminoglycoside phosphotransferase
dsDNA	double-stranded DNA	NGS	next-generation sequencing
DSIF	DRB sensitivity inducing factor	nt	nucleotide
dTSR	divergent TSR	NuA4	nucleosome acetyltransferase of H4
E	glutamic acid	oligo	oligonucleotide
<i>et al.</i>	<i>et alia</i>	ORF	open reading frame
F	phenylalanine	ψ221	VSG pseudogene of active 221 BES
FCS	fetal calf serum	P	proline
Fluc	firefly luciferase	PIC	pre-initiation complex
G (nucleobase)	guanine	P-TEFb	positive transcription elongation factor b
G (amino acid)	glycine	PAC	puromycin N-acetyltransferase
G418	neomycin	PAF	polymerase II-associated factor
GEO	Gene Expression Omnibus	PAGE	polyacrylamide gelelektrophoresis
GTF	general transcription factor		
H	histidine		
H2A	histone H2A		

PBS	phosphate buffered saline	V	volume
PCR	polymerase chain reaction	VSG	variant surface glycoprotein
Phleo	phleomycin	W	tryptophan
Pol	polymerase	WHO	World Health Organization
polyT	polythymine	ws	window size
polyY	polypyrimidine	Y (nucleobase)	pyrimidine, IUPAC code
pre-mRNA	precursor-mRNA	Y (amino acid)	tyrosine
PTM	post-translational modification	V	valine
PTU	polycistronic transcription unit		
Puro	puromycin		
Q	glutamine		
qPCR	quantitative PCR		
R	arginine		
RIPA	Radio immunoprecipitation assay buffer		
Rluc	renilla luciferase		
RNA	ribonucleic acid		
rpm	rounds per minute		
rRNA	ribosomal RNA		
RSC	chromatin structure remodeling		
RT	room temperature		
S	serine		
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>		
SAGA	Spt-Ada-Gcn5 Acetyltransferase		
SANT	Swi3, Ada2, N-Cor, and TFIIB		
SAS	splice acceptor site		
SC	subcutaneous		
SD	standard deviation		
SDS (splicing)	splice donor site		
SDS (chemical)	sodium dodecyl sulfate		
Ser	serine		
Set1	Su(var)3-9, Enhancer-of-zeste, Trithorax 1		
SL	spliced leader		
SLIDE	SANT-like ISWI domain		
SM	single marker		
snRNP	small nuclear ribonucleicprotein		
Sth1	RSC chromatin remodeling complex ATPase subunit		
ss	step size		
ssDNA	single-stranded DNA		
SWI/SNF	switching defective/sucrose nonfermenting		
T (nucleobase)	thymine		
T (amino acid)	threonine		
T7RNAP	T7 RNA polymerase		
TBP	TATA-box binding protein		
<i>T. brucei</i>	<i>Trypanosoma brucei</i>		
TERT	telomerase reverse transcriptase		
Tet	tetracycline		
TetO	tetracycline operator		
TetR	tetracycline repressor		
TEX	terminator exonuclease		
TF	transcription factor		
TLCK	Tosyl-L-lysyl-chloromethane hydrochloride		
TRF4	TBP-related protein 4		
tRNA	transfer-RNA		
TSR	transcription start region		
TSS	transcription start site		
U	unit		
UAS	upstream activating sequece		
UHR	upstream homology region		
URS	upstream repressing sequence		
UTR	untranslated region		

List of figures

Figure 1.1 Promoter elements.....	19
Figure 1.2 Fully assembled RNA polymerase II pre-initiation complex (PIC).....	20
Figure 1.3 Chromatin structure.....	22
Figure 1.4 Nucleosomal organization around the TSS in <i>Saccharomyces cerevisiae</i>	25
Figure 1.5 Sequence preferences during nucleosome formation.	27
Figure 1.6 Mechanism of histone exchange.....	33
Figure 1.7 Mechanism of RNA splicing.....	35
Figure 1.8 Nucleosome occupancy across the 3' splice acceptor site (SAS).....	36
Figure 1.9 Genes are organized in PTUs and <i>trans</i> -spliced in <i>T. brucei</i>	39
Figure 1.10 Epigenetic marks at PTUs.....	42
Figure 2.1 MNase activity varies with concentration and substrate.....	71
Figure 2.2 NGS library construction.....	72
Figure 2.3 The number of required reads depends on the scope of the experiment.....	73
Figure 3.1 RPB9 is enriched at the 5'-end of H2A.Z enrichment.....	80
Figure 3.2 Identification of transcription initiation sites by mapping small primary transcripts.	81
Figure 3.3 Establishment of a high-resolution MNase-ChIP-seq protocol for <i>T. brucei</i>	83
Figure 3.4 Fragment size distribution and dinucleotide frequencies upon MNase-ChIP-seq. ...	85
Figure 3.5 TSRs show increased MNase sensitivity.....	86
Figure 4.1 Examined TSR DNA sequences.....	89
Figure 4.2 Approach to target TSR DNA sequences to a non-transcribed locus.	90
Figure 4.3 TSR DNA sequences are capable to mediate transcription initiation dependent on the genomic location.....	91
Figure 4.4 DNA elements are distributed across TSRs and provide directionality to transcription.	93
Figure 4.5 GT-rich sequence elements on the coding strand mediate transcription initiation... 95	95
Figure 4.6 Comparison of endogenous expression levels.....	96
Figure 4.7 Influence of GT-rich sequence insertion on the transcription of flanking PTUs.....	97
Figure 4.8 H2A.Z enrichment and luciferase activity increase over time.....	98
Figure 5.1 Most genes within PTUs are preceded by an NDR.	101
Figure 5.2 Nucleosome depletion correlates with the level of gene expression.....	102
Figure 5.3 Composition of polyY tract affects gene expression.....	103
Figure 5.4 Composition of polyY tract affects nucleosome positioning.....	104

Figure 6.1 The structure of Z-DNA..... 110
Appendix Figure 7.1 Generation of α H2A.Z and α H2B.V..... 117
Appendix Figure 7.2 Nucleosome depletion correlates with the level of gene expression. 118

List of tables

Table 1.1 Promoter DNA elements.	18
Table 1.2 Selection of organisms for which genome-wide nucleosome positioning maps have been generated.	24
Table 1.3 ATP-dependent chromatin remodeler complexes in <i>S. cerevisiae</i>	28
Table 2.1 List of parental constructs used in this study.	48
Table 2.2 List of constructs generated in this study.	53
Table 2.3 List of oligos used in this study.	55
Table 2.4 List of gBlocks used in this study.	60
Table 2.5 List of parental <i>T. brucei</i> cell lines used in this study.	61
Table 2.6 List of transgenic <i>T. brucei</i> cell lines generated in this study.	62
Table 2.7 Software used in this study.	77
Appendix Table 7.1 List of RNA pol II transcription initiation sites.	119
Appendix Table 7.2 List of 10mers enriched to at least 6-fold on coding strand compared to the noncoding strand across TSRs.	121
Appendix Table 7.3 Information about sequencing data discussed in this study.	124
Appendix Table 7.4 Information about the processing of the sequencing data discussed in this study.	125

Table of contents

Affidavit.....	I
Acknowledgements.....	II
Summary.....	III
Zusammenfassung.....	IV
Abbreviation index.....	VI
List of figures.....	VIII
List of tables.....	X
Table of contents.....	XI

Introduction: Regulation of transcription initiation in eukaryotes 15

1.1 The pre-initiation complex.....	17
1.1.1 Regulatory promoter elements.....	17
1.1.2 Assembly of the RNA pol II pre-initiation complex.....	19
1.2 Chromatin structure.....	20
1.2.1 DNA is packaged into chromatin.....	21
1.2.2 Methods to study nucleosome positioning.....	22
1.2.3 Chromatin structure around promoters.....	23
1.2.4 Determinants of nucleosome formation and positioning across the genome.....	25
1.2.4.1 DNA sequence.....	26
1.2.4.2 Chromatin remodelers.....	28
1.2.4.3 Transcription factors and active transcription.....	30
1.2.4.4 Suggested integrative models.....	30
1.2.5 Chromatin regulates post-transcription-initiation events.....	31
1.2.5.1 Promoter-proximal RNA pol II pausing.....	31
1.2.5.2 Transcription elongation.....	32
1.2.5.3 Splicing.....	34
1.3 Focused and dispersed transcription initiation.....	36
1.4 <i>Trypanosoma brucei</i>	37
1.4.1 General overview.....	37
1.4.2 Gene expression in <i>T. brucei</i>	38
1.4.2.1 Genes are organized in polycistronic transcription units.....	38
1.4.2.2 <i>trans</i> -splicing.....	39
1.4.2.3 Transcription initiation in <i>T. brucei</i>	40
1.5 Aim of the study.....	42

Materials and methods 44

2.1 Molecular cloning methods.....	45
2.1.1 Polymerase chain reaction (PCR).....	45

2.1.2	Restriction digest	45
2.1.3	InFusion and transformation	46
2.1.4	Ligation and transformation	46
2.1.5	Plasmid isolation	47
2.1.6	Sanger sequencing	47
2.1.7	Bacterial stock preparation	47
2.1.8	EtOH precipitation	47
2.2	Generation of constructs.....	48
2.2.1	Generation of pPOTv3_TY-RPB9_Phleo/Puro.....	48
2.2.2	Generation of pLEW111_TY1-H2A.Z.....	49
2.2.3	Generation of pyrFEKO-HYG/PUR_H2A.Z.....	49
2.2.4	Generation of TSR translocation constructs.....	50
2.2.5	Generation GT-rich promoter constructs.....	51
2.2.6	Generation of pCW37.....	52
2.2.7	Generation of polyY constructs.....	52
2.3	Trypanosome cell culture and analysis	61
2.3.1	Trypanosome growth.....	61
2.3.2	Cryo stock preparation and reconstitution.....	64
2.3.3	Stable transfection of <i>T. brucei</i> cells	64
2.3.4	Transient transfection of <i>T. brucei</i> cells.....	64
2.3.5	Isolation of genomic DNA	65
2.3.6	Isolation of RNA, cDNA synthesis and qPCR analysis.....	65
2.3.7	Dual-Luciferase assay.....	65
2.3.8	Fluorescence microscopy.....	66
2.4	Biochemical methods	67
2.4.1	Western blot.....	67
2.4.2	Antibody production.....	67
2.4.3	Antibody affinity purification	68
2.5	Next-generation sequencing methods	69
2.5.1	MNase-ChIP-seq.....	69
2.5.2	Library construction.....	71
2.5.3	Library quantification and sequencing	72
2.5.4	Computational analysis	74
2.5.5	Ty1-RPB9-ChIP-seq	76
2.5.6	5' PPP-RNA-seq.....	76

2.6 Data generated in this study and source code availability.....	77
2.7 Software	77
Characterization of RNA pol II transcription start regions	78
3.1 The RNA pol II subunit RPB9 is enriched at the 5'-end of TSRs.....	79
3.2 Transcription initiates ~200 bp upstream of RPB9 enrichment	80
3.3 Chromatin structure around transcription start regions	82
3.3.1 MNase-ChIP-seq - A high-resolution method to investigate chromatin accessibility by mapping nucleosome positioning	82
3.3.2 Sites enriched in H2A.Z show increased sensitivity to MNase	84
3.4 Concluding remarks.....	86
GT-rich DNA sequences can initiate transcription and induce targeted H2A.Z deposition	87
4.1 TSR DNA sequences are capable to initiate transcription.....	88
4.1.1 Insertion of TSR DNA sequences in transcriptional silent locus	89
4.1.2 TSR DNA sequence-mediated transcription initiation is dependent on the genomic locus	90
4.1.3 The transcription-mediating sequence element is distributed across TSRs and directs transcription	92
4.2 GT-rich promoter elements can trigger transcription initiation	93
4.3 GT-rich promoter elements promote targeted H2A.Z deposition.....	97
4.4 Concluding remarks.....	98
Nucleosome depleted regions at exon boundaries are affected by the DNA sequence	99
5.1 Exon boundaries rather than TSRs contain well-defined NDRs.....	100
5.2 Composition of the polyY tract affects gene expression and nucleosome positioning	102
5.3 Concluding remarks.....	104
Discussion.....	106
6.1 Transcription initiates at the 5'-end of TSRs	107
6.2 RNA pol II transcription initiation is DNA sequence-mediated	108
6.3 GT-rich promoter elements contribute to targeted H2A.Z deposition.....	111

6.4 NDRs regulate gene expression post-transcriptionally and are affected by the DNA sequence	113
6.5 Conclusion	114
Appendix	116
7.1 Appendix Figures.....	117
7.2 Appendix Tables.....	119
References.....	CXXVI
Curriculum vitae	CXXXVII
List of publications.....	CXXXVIII
Attended conferences and courses.....	CXXXIX

1

Introduction:

Regulation of transcription initiation in eukaryotes

1.1	The pre-initiation complex	17
1.2	Chromatin structure	20
1.3	Focused and dispersed transcription initiation.....	36
1.4	<i>Trypanosoma brucei</i>	37
1.5	Aim of the study	42

Transcription is a fundamental process conducted in all living cells and essential to transfer the genomic information into RNA, which is then translated into an amino acid sequence to produce proteins (Crick, 1970). Whereas the amount of genomic information is constant, the amount of protein produced from different genes can vary greatly. This is facilitated and regulated by the mechanism of gene expression, which involves the action of several cellular processes such as transcription, splicing and translation (Jocelyn *et al.*, 2011). Thus, the regulation of these processes is of central importance for life.

The genomic information is encoded in the sequence of the DNA, the basal molecule involved in gene expression. Its sequence is read by an RNA polymerase (RNA pol) and converted into RNA, a process termed transcription. The nascent RNA contains intron sequences, which are removed by splicing to produce mature messenger RNA (mRNA). Splicing can occur during transcription, while the remaining RNA is still transcribed and after transcription has finished. There are three distinct RNA polymerases in eukaryotes that transcribe different classes of genes. While RNA pol I transcribes ribosomal RNAs (rRNAs) and RNA pol III transfer RNAs (tRNAs), RNA pol II transcribes protein-coding genes, which yield mRNAs. Although, polymerases are recruited to different classes of promoters they share some common features (Geoffrey, 2000). The first crucial step during transcription is the recruitment of the RNA pol, which leads to subsequent transcription initiation. This is a highly regulated process that requires the assembly of an RNA pol pre-initiation complex (PIC), which results from a cascade of protein-binding events to the promoter (Venters and Pugh, 2009a).

Transcription is a DNA-templated process, meaning it relies on the contact between the DNA double helix and DNA-binding proteins and is thereby dependent on the accessibility of the DNA. In eukaryotic genomes, DNA is packaged into chromatin, a structure that is composed of nucleosomes, in which DNA is wrapped around histone proteins (Luger *et al.*, 1997). The tightness of the wrapping and the distance between nucleosomes dictate the accessibility of the DNA and can be altered by several factors to generate e.g. nucleosome depleted regions (NDRs). While promoter regions are depleted of nucleosomes, gene bodies are protected by nucleosome formation (reviewed in Hughes and Rando, 2014). Thus, the PIC may assemble within the promoter region, but the progression of transcription is inhibited by the presence of nucleosomes. Hence, chromatin structure provides an additional level of transcription initiation regulation. In addition, splice sites are depleted of nucleosomes, which might facilitate a faster transcription within these regions and thereby regulate co-transcriptional splicing (reviewed in Naftelberg *et al.*, 2015).

In both mechanisms, PIC assembly and the establishment of chromatin structures, the DNA sequence plays a fundamental role. It may contain DNA sequence elements that recruit

subunits of the RNA pol complex and therefore define promoter regions. Furthermore, the DNA sequence plays a central role in the establishment of the chromatin structure, especially around promoters where it often promotes the formation of nucleosome depleted regions, leaving the DNA accessible for proteins (Raisner *et al.*, 2005).

In the following chapters, I will summarize the mechanisms of RNA pol II PIC assembly and chromatin formation with respect to the role of the DNA sequence. In addition, I will focus on the regulatory impact of chromatin on transcription and introduce two distinct classes of transcription initiation, focused and dispersed. Finally, I will describe how the eukaryotic parasite *Trypanosoma brucei* appears to survive without any transcriptional control.

1.1 The pre-initiation complex

Transcription initiation requires the binding of RNA pol II to the promoter yet RNA pol II itself does not recognize promoter sequence motifs (Juven-Gershon and Kadonaga, 2010). Thus, the polymerase depends on other proteins to be part of the pre-initiation complex, which bind to regulatory sequence elements within promoters and recruit RNA pol II to the transcription initiation site (reviewed in Venters and Pugh, 2009a; Juven-Gershon and Kadonaga, 2010; Roy and Singer, 2015). In this section, regulatory DNA elements within the promoter will be described followed by an overview about the assembly of the RNA pol II PIC mediated by these regulatory elements.

1.1.1 Regulatory promoter elements

Regulatory promoter elements can be located within a core promoter, which is a 50-100 bp sequence where the transcription machinery assembles and transcription initiates. So-called upstream elements can be located in yeast 100-500 bp and in metazoans up to several thousands of bp upstream of the core promoter (Harbison *et al.*, 2004; Venters and Pugh, 2009a). These sequence elements are bound by *trans*-acting sequence-specific factors, such as the general transcription factors (GTFs), which recruit the transcription machinery including RNA pol II. Core promoter and upstream elements are listed in Table 1.1 and illustrated in Figure 1.1.

Table 1.1 Promoter DNA elements.

Adapted from (Venters and Pugh, 2009a; Juven-Gershon and Kadonaga, 2010; Roy and Singer, 2015). TSS, transcription start site.

DNA element	Name	Consensus sequence (IUPAC)	Distance from TSS	Species
<i>Core promoter elements</i>				
TATA	TATA-box	TATAWAAR	-25 bp	Yeast, metazoans
Inr	Initiator	YYANWYY	-2 to +4 bp, spans TSS	Metazoans
BRE	TFIIB recognition element	SSRCGCC (BREu), RTWKKKK (BREd)	-35 bp (BREu), -20 bp (BREd)	Metazoans
DPE	Downstream promoter element	RGWCGTG	+30 bp	Metazoans
DCE	Downstream core element	CTTC, CTGT, AGC	+9 bp, +18 bp, +32 bp	Metazoans
MTE	Motif ten element	CSARCSAACG	+23 bp	Metazoans
<i>Upstream elements</i>				
UAS	Upstream activating sequence	-	-several hundreds of bp	Yeast
Enhancer	-	-	-several thousands of bp	Metazoans
URS	Upstream repressing sequence	-	-several hundreds of bp	Yeast

The most common core promoter motifs are the TATA-box and the initiator element (Inr), which occur either together or separately (Yang *et al.*, 2007; Frith *et al.*, 2008). Both are binding sites for the GTF TFIID. The TATA-box is specifically bound by the TFIID subunit TATA-binding protein (TBP), which initiates the assembly of the RNA pol II pre-initiation complex (Roeder, 1996). Twenty percent of promoters in yeast and 10-15% of mammalian promoters contain a TATA-box (Kim *et al.*, 2005; Carninci *et al.*, 2006; Cooper *et al.*, 2006), while 70% and 46% of *Drosophila* and human promoters contain an Inr element (Smale and Kadonaga, 2003; Yang *et al.*, 2007). Both elements appear to serve distinct functions: genes controlled by a TATA-box-containing promoter are mostly tissue-specific and need a tight regulation (Lenhard *et al.*, 2012), whereas Inr-containing promoters control ubiquitously expressed genes (Gershenson and Ioshikhes, 2005; Sandelin *et al.*, 2007; Lenhard *et al.*, 2012). The TFIIB recognition element (BRE) is composed of two distinct motifs flanking the TATA-box (BREu and BREd) and serves as the binding site for the GTF TFIIB. The downstream promoter element (DPE) is

present in mammalian and *Drosophila* Inr promoters (Smale and Kadonaga, 2003). It serves as the downstream core element (DCE) while the Inr-dependent motif ten element (MTE) serves as a binding site for components of the TFIID complex (Lewis *et al.*, 2000; Juven-Gershon *et al.*, 2008; Roy and Singer, 2015). Elements located further upstream of the core promoter are called upstream activating/repressing sequences (UAS/URS) in yeast and enhancers in metazoans. They are bound by *trans*-activating factors which results in the case of UAS and enhancers in activation of transcription or in case of URS in decreased transcriptional activity (Venters and Pugh, 2009a).

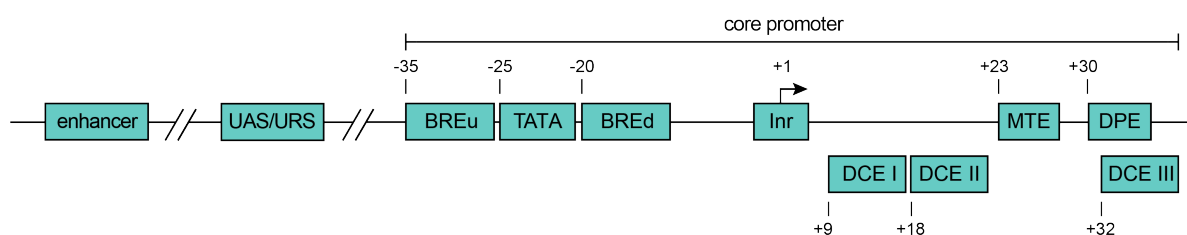


Figure 1.1 Promoter elements.

Promoter elements found in yeast and metazoans relative to the TSS (+1). UAS, upstream activating sequence; URS, upstream repressing sequence; BREu, upstream TFIIB recognition element; TATA, TATA-box; BREd, downstream TFIIB recognition element; Inr, Initiator; DCE, downstream core element; MTE, motif ten element; DPE, downstream promoter element.

1.1.2 Assembly of the RNA pol II pre-initiation complex

The PIC is composed of GTFs and RNA pol II and is assembled at promoters. The detailed composition of subunits may vary between yeast and mammals. Common subunits encompass the following proteins: TBP (TATA-box binding protein), and the GTFs TFIIB, TFIIF, TFIIE and TFIIH (He *et al.*, 2013; Murakami *et al.*, 2013). Since those have been identified in the context of TATA-box-containing promoters the following characteristics are restricted to TATA-box-containing promoters. TBP is usually part of the TFIID complex and binds to the TATA-box located within an accessible region of promoters. The binding results in a bending of the DNA (Figure 1.2; Kim *et al.*, 1993). A second factor, TFIIB, binds to the TATA-box flanking BREs (Lagrange *et al.*, 1998; Deng and Roberts, 2005). It additionally interacts with TBP resulting in an interaction of a TFIIB domain with RNA pol II (Sainsbury *et al.*, 2013) and thereby in the recruitment of RNA pol II to the PIC. RNA pol II is not able to melt the DNA to separate the template from the non-template strand unlike other polymerases. It is aided by TFIIH, which is the only GTF exhibiting enzymatic activity by containing two helicases Ssl2p and Rad3p and a kinase Kin28p (in yeast; Tirode *et al.*, 1999). The site of interaction of TFIIH is not clear yet, but it is suggested that it interacts with the DNA

downstream of the complex. TFIIIE facilitates the binding of TFIIH and also the melting of the promoter (reviewed in Luse, 2014). For TFIIIF it is suggested that it facilitates the loading of RNA pol II into the PIC and stabilizes TFIIIB within the PIC (reviewed in Luse, 2012). Finally, the transition from transcription initiation to elongation is facilitated by phosphorylation of serine 5 within the YSPTSPS heptapeptide repeats in the C-terminal domain (CTD) of RNA pol II mediated by the kinase activity of the TFIIH subunit (Valay *et al.*, 1995).

The assembly of the PIC within promoters is not the sole regulator of transcription initiation. In most eukaryotic organisms, the genomic DNA is packaged into chromatin, which makes the DNA less accessible for protein-binding events. The following section will describe how the chromatin status is regulated to allow DNA-templated processes such as PIC assembly.

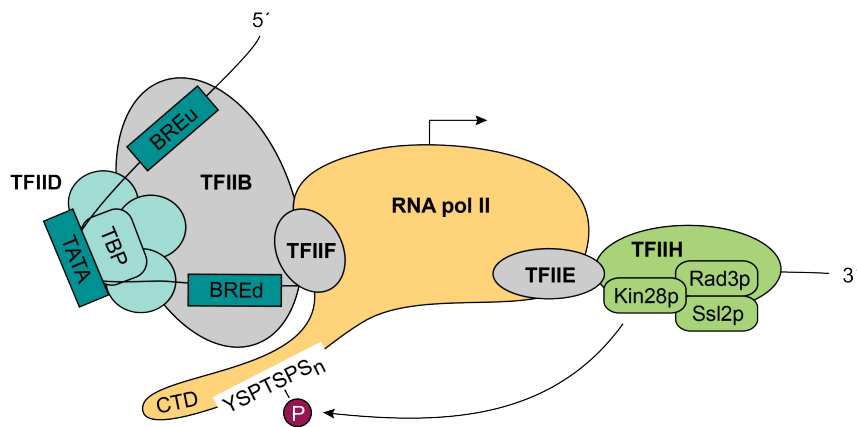


Figure 1.2 Fully assembled RNA polymerase II pre-initiation complex (PIC).

Schematic illustration of the RNA polymerase II (RNA pol II) pre-initiation complex on a TATA-box-containing promoter. The TATA-box is bound by the TATA-box binding protein (TBP), which is part of the general transcription factor (GTF) TFIID. TFIIIB binds to the TATA-box flanking TFIIIB recognition elements (BREu, BREd) and TBP and thereby recruits RNA pol II to the PIC. TFIIIF aids the loading of RNA pol II and stabilizes TFIIIB. TFIIIE facilitates the binding of TFIIH that contains 2 helicases (Rad3p, Ssl2P) and a kinase (Kin28p). Kin28p phosphorylated Ser5 within the heptapeptide repeat in the C-terminal domain (CTD) of RNA pol II and RNA pol II initiates transcription.

1.2 Chromatin structure

The chromatin structure influences all steps of transcription, namely: initiation, promoter escape, elongation and termination (Venkatesh and Workman, 2015). After a brief introduction into chromatin biology the following sections will focus on the detailed nucleosome positioning as nucleosomes are the basic subunits of chromatin. First, I will give an overview about state-of-the-art methods to study nucleosome positioning, subsequent sections will focus on the chromatin structure around promoters and describe determinants of nucleosome positioning. Furthermore, the role of chromatin in the regulation of post-transcription-initiation events, e.g.

promoter proximal RNA pol II pausing, transcription elongation and co-transcriptional splicing will be addressed.

1.2.1 DNA is packaged into chromatin

Eukaryotic DNA is packaged into chromatin, which consists of repeating nucleoprotein complexes called nucleosomes (Figure 1.3; Hewish and Burgoyne, 1973; Kornberg and Thomas, 1974; Olins and Olins, 1974). One nucleosome consists of 147 bp of DNA wrapped 1.65 times around a histone octamer composed of two copies of each core histone H2A, H2B, H3 and H4 (Luger *et al.*, 1997). Nucleosomes are separated by the linker DNA, which contains 0-80 bp and varies among species and cell types (Prunell and Kornberg, 1982). One molecule of the linker histone H1 is bound to the entry and exit sites of nucleosomal DNA and completes the second turn of wrapped DNA (Allan *et al.*, 1980). This subunit is called chromatosome and consists of 167 bp of nucleosomal DNA (Widom, 1989, 1998). In the chromatosome conformation, the so-called 10 nm fiber, the DNA is condensed about 6-fold. The 10 nm fiber can be further condensed to the 30 nm fiber, a structure that still needs to be resolved and has been questioned in some publications (Lieberman-Aiden *et al.*, 2009; Fussner *et al.*, 2012; Gan *et al.*, 2013). Originally, it was thought that the sole purpose of packaging of DNA into chromatin was to fit the large amount of eukaryotic DNA (e.g. 2 m for all 46 human chromosomes) into the nucleus of approx. 6 μm in diameter (Bruce *et al.*, 2014). Yet, within the last 20 years it became clear that the degree of chromatin compaction is an important regulatory mechanism for DNA templated processes such as recombination, DNA repair, replication and gene expression (Hughes and Rando, 2014). Especially the positioning of nucleosomes along the DNA is of central interest since it affects the accessibility of the DNA to binding proteins. Thus, a range of methods have been developed to study nucleosome formation and positioning.

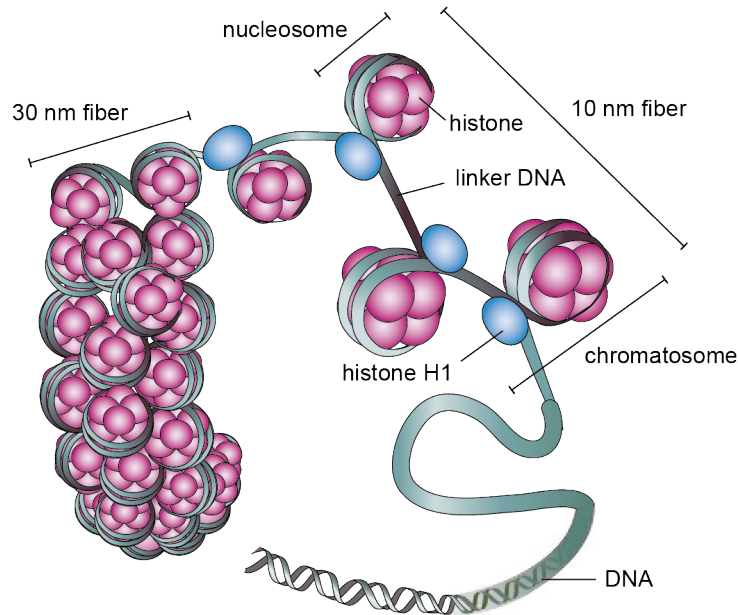


Figure 1.3 Chromatin structure.

A nucleosome is the basic unit of chromatin and consists of 147 bp of DNA wrapped around a histone octamer. Histone H1 binds to the DNA entry and exits sites at nucleosomes and the complex is called chromatosome. The 10 nm fiber, in which nucleosomes are located next to each other can be further condensed to a 30 nm fiber (modified from Figueiredo *et al.*, 2009).

1.2.2 Methods to study nucleosome positioning

A nucleosome's position relative to a given DNA locus can be studied by identifying the part of the DNA that is associated with the histone octamer (nucleosomal DNA) and is referred to as translational nucleosome positioning (Satchwell *et al.*, 1986). Thereby, nucleosome positioning can be studied by using methods that map protein-DNA interactions at high resolution, such as chromatin immunoprecipitation (ChIP) combined with next-generation sequencing (ChIP-seq). Within the last 10 years ChIP-seq has become an indispensable tool for mapping protein-DNA interactions (Johnson *et al.*, 2007) and the determination of genome-wide nucleosome positioning (Albert *et al.*, 2007). During ChIP-seq, DNA-protein interactions are preserved by *in vivo* formaldehyde crosslinking and the chromatin is fragmented by sonication. The protein-DNA complex of interest is pulled-down with a specific antibody and crosslinks are reversed for subsequent DNA analysis. The immunoprecipitated DNA is sequenced and mapped to a reference genome revealing the profile of protein-DNA binding events, e.g. nucleosomes. The resolution of ChIP-seq is amongst others influenced by the size of the fragmented DNA (Fan *et al.*, 2008). Typically, sonication-based fragmentation of chromatin yields fragments between 200 and 500 bp (Mahony and Pugh, 2015). Thus, ChIP-seq is not perfectly suited for the precise mapping of nucleosomes (Albert *et al.*, 2007).

One possibility to improve spatial resolution experimentally is the implementation of an enzymatic cleavage step after immunoprecipitation, as it is performed in ChIP-exo assays (Rhee and Pugh, 2011, 2012). Here, the lambda exonuclease is used to digest the immunoprecipitated DNA in 5'-3' direction. This digestion removes one strand of the DNA until it is protected by the bound protein and also reduces background by the removal of non-crosslinked DNA. After crosslinks are reversed, the digested strand is synthesized by primer extension (the undigested strands were ligated to another adapter than the newly synthesized) and sequenced from one end. Although this approach adds more resolution to the sequencing data, more input material compared to sonication-based ChIP-seq assays is required due to additional washing steps.

Alternatively, the enzymatic fragmentation step can be performed before the immunoprecipitation step as it is the case during MNase-ChIP-seq. Here, the fragmentation of chromatin is facilitated by treatment with micrococcal nuclease (MNase). MNase is a non-specific endonuclease derived from *S. aureus* with a preference for AT-dinucleotides that digests dsDNA, ssDNA as well as RNA. It digests unprotected DNA, e.g. linker DNA between nucleosomes, leaving a 'footprint' of the bound histones. Thus, it is well suited to generate accurate genome-wide nucleosome positioning maps at high resolution (Albert *et al.*, 2007). In comparison to sonication-based approaches, MNase-ChIP-seq provides a high reproducibility due to the use of an enzymatic cleavage instead of random sheering of the chromatin. In addition, fewer washing steps are needed, which reduced the hands-on time and the number of cells needed for each assay.

1.2.3 Chromatin structure around promoters

Conducting the above-mentioned methods, it was possible to generate genome-wide nucleosome positioning maps. The first genome-wide nucleosome positioning analysis was performed in *S. cerevisiae* and to date high-resolution nucleosome positioning maps have been generated for more than 30 organisms and multiple cell types (a selection is listed in Table 1.2).

Table 1.2 Selection of organisms for which genome-wide nucleosome positioning maps have been generated.

Organism	Reference
<i>Saccharomyces cerevisiae</i>	(Yuan <i>et al.</i> , 2005)
<i>Schizosaccharomyces pombe</i>	(Lantermann <i>et al.</i> , 2009)
<i>Dictyostelium discoideum</i>	(Chang <i>et al.</i> , 2012)
<i>Aspergillus fumigatus</i>	(Nishida <i>et al.</i> , 2009)
<i>Arabidopsis thaliana</i>	(Chodavarapu <i>et al.</i> , 2010)
<i>Caenorhabditis elegans</i>	(Valouev <i>et al.</i> , 2008)
<i>Plasmodium falciparum</i>	(Westenberger <i>et al.</i> , 2009)
<i>Drosophila melanogaster</i>	(Mavrich <i>et al.</i> , 2008)
<i>Oryzias latipes</i>	(Nahkuri <i>et al.</i> , 2009)
<i>Mus musculus</i>	liver cells (Li <i>et al.</i> , 2011), embryonic stem cells (Teif <i>et al.</i> , 2012)
<i>Homo sapiens</i>	CD4 ⁺ T cells (Schones <i>et al.</i> , 2008), embryonic stem cells (Yazdi <i>et al.</i> , 2015)

All studies revealed a common and thereby evolutionary conserved pattern: nucleosomes occupy preferred positions in genes and non-gene regions and they are depleted at most of the promoters and some enhancers and terminators, generating so-called nucleosome depleted regions (NDRs). In general, the size of NDRs varies with the class of promoters (for more information see chapter 1.3) and the phase of the cell cycle (Kelly *et al.*, 2010; Nekrasov *et al.*, 2012). In yeast, NDRs have an average size of 150 bp (Jiang and Pugh, 2009), are enriched in transcription factor (TF) binding sites (Lee *et al.*, 2007b; Ozonov and van Nimwegen, 2013), TATA boxes or TATA-like elements (Basehoar *et al.*, 2004; Rhee and Pugh, 2012) and are accessible to the pre-initiation complex of RNA pol II (Morse, 2007). NDRs are flanked by at least one nucleosome containing the H2A histone variant H2A.Z and the respective first nucleosome is designated as -1 (upstream of NDR) and +1 nucleosome (downstream of NDR). Both nucleosomes and downstream located nucleosomes are well-positioned, meaning their translational positioning is the same in each cell of the population generating a phased nucleosome array (Figure 1.4). In yeast, the +1 nucleosome often occludes the transcription start site (TSS; Albert *et al.*, 2007), whereas it is positioned

downstream of the TSS in metazoans (Mavrigh *et al.*, 2008; Schones *et al.*, 2008). The extent of the phasing decreases gradually towards the 3'-end of the coding region and it seems to be dependent on transcriptional activity, since it is not observed upstream of the NDR and at repressed genes (Schones *et al.*, 2008; Lantermann *et al.*, 2010; Valouev *et al.*, 2011). H2A.Z-containing nucleosomes have been shown to be less stable, facilitating an easier eviction of the nucleosome during transcription initiation (Suto *et al.*, 2000; Abbott *et al.*, 2001; Zhang *et al.*, 2005; Jin and Felsenfeld, 2007; Siegel *et al.*, 2009). Furthermore, nucleosomes downstream of the promoter are characterized by histones with specific post-translational modifications, such as acetylation on lysine 16 of the H4 tail (Kimura *et al.*, 2002; Suka *et al.*, 2002) and methylation of lysines 4 and 79 of H3 (van Leeuwen *et al.*, 2002; Ng *et al.*, 2003a; Santos-Rosa *et al.*, 2004).

This characteristic chromatin structure around promoters is evolutionarily conserved and emerges genome-wide in the individual organisms. The following sections will describe the underlying mechanisms dictating which part of the DNA sequence is included in the nucleosome and the composition of the histone octamer.

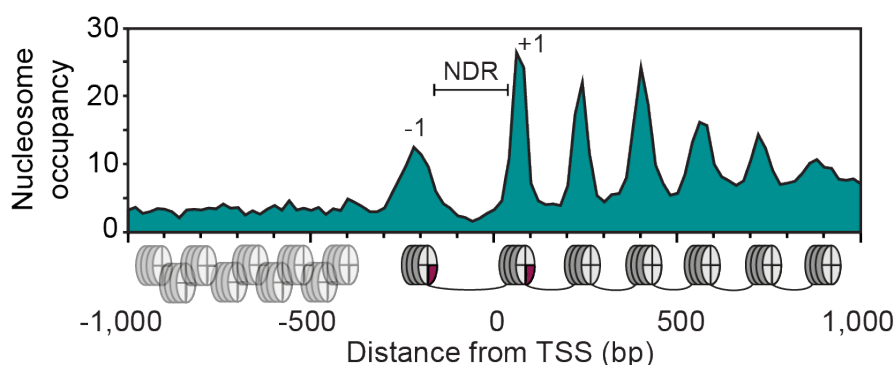


Figure 1.4 Nucleosomal organization around the TSS in *Saccharomyces cerevisiae*.

Positioning of nucleosomes across the TSS of all coding genes within the *S. cerevisiae* genome (modified from Mavrigh *et al.*, 2008). Nucleosome occupancy indicates how strong a certain DNA sequence is occupied by a nucleosome. The nucleosomes flanking the nucleosome depleted region (NDR) are marked with -1 and +1, resp. and contain H2A.Z (shown in pink).

1.2.4 Determinants of nucleosome formation and positioning across the genome

Intense studies within the last 30 years revealed that the translational nucleosome positioning is largely but not exclusively determined by the DNA sequence. The final *in vivo* nucleosome positioning is achieved by the interplay of factors and mechanisms that 'read' the DNA sequence, such as histone-DNA interactions, chromatin remodelers, TFs and active transcription. But the relative importance of each is not fully understood yet. From a

nucleosome point of view, the mentioned factors can be divided into two distinct classes: *cis*- (DNA sequence, histones) and *trans*-acting (chromatin remodelers, transcription TFs and active transcription) factors (Radman-Livaja and Rando, 2010). If not mentioned otherwise, the following findings rely on studies performed in *S. cerevisiae*, since many of the basic mechanisms are highly conserved among eukaryotes and most of the research has been done in this organism.

1.2.4.1 DNA sequence

Regarding the extent to which the DNA sequence contributes to nucleosome positioning, the following two distinct hypotheses have been put forward: i) each individual nucleosome position is determined by a 'genomic code for nucleosome positioning' (Segal *et al.*, 2006; Kaplan *et al.*, 2009), where nucleosome positioning is determined by sequence preferences of the histone octamer and by the bending ability of the DNA strand and ii) only a few positions are sequence-determined and the majority of nucleosomes is positioned stochastically relative to the positioned nucleosomes or e.g. a TF binding site acting as a barrier. This mechanism is referred to as 'statistical positioning' (Kornberg and Stryer, 1988; Möbius and Gerland, 2010).

Histones bind to DNA genome-wide, thus they do not possess sequence-specific DNA binding domains. The interactions between histones and DNA rely on the penetration of the amino acid residues of the histones into the minor groove of the phosphodiester backbone of the DNA (Luger *et al.*, 1997). Early *in vitro* studies showed that nucleosome formation is highly dependent on the ability of DNA to bend around the histone octamer in a DNA sequence dependent manner and they identified sequences favored and disfavored for nucleosome formation. These studies revealed a preferred rotational positioning, e.g. the orientation of the DNA helix on the histone surface, of AT-dinucleotides in a 10 bp periodicity (1 helical turn) in the minor groove facing the histone octamer and of GC-dinucleotides facing outward (Figure 1.5; Drew and Travers, 1985; Satchwell *et al.*, 1986). This behavior could be confirmed by the generation of a genome-wide nucleosome positioning map (Brogaard *et al.*, 2012) and is due to the fact that the minor groove of AT-rich sequences is narrower compared to those of GC-rich sequences (Olson and Zhurkin, 2011). In addition, homopolymeric poly(dA:dT) and poly(dG:dC) sequences have been identified to be disfavored by nucleosome formation because they are intrinsically rigid (McCall *et al.*, 1985; Nelson *et al.*, 1987; Suter *et al.*, 2000; Segal and Widom, 2009; Tsankov *et al.*, 2011). While poly(dA:dT) tracts are abundant in eukaryotic genomes (Dechering *et al.*, 1998), its prevalence at promoter regions varies among species. Whereas poly(dA:dT) occurrence is very common in *S. cerevisiae* (Yuan *et al.*, 2005; Lee *et al.*, 2007b) it is rather rare in higher eukaryotes (Lantermann *et al.*, 2010; Tsankov

et al., 2010, 2011). In yeast, it has been shown, that poly(dA:dT) help positioning the +1 nucleosome (Raisner *et al.*, 2005; Zhang *et al.*, 2009) and the introduction of a poly(dA:dT) tract can generate an NDR (Small *et al.*, 2014). Widom and colleagues identified a sequence from a pool of synthetic DNA sequences that mediates strong nucleosome positioning following the same rules. It exhibits an even stronger histone affinity than naturally occurring sequences and was named the Widom601 sequence (Lowary and Widom, 1998). The role of the DNA sequence during genome-wide nucleosome positioning *in vivo* has been addressed by *in vitro* experiments, in which nucleosomes are assembled by salt gradient dialysis using purified histones and genomic DNA. Although the generated *in vitro* nucleosome pattern at non-promoter intergenic regions is notably similar compared to the *in vivo* pattern, it significantly differs around the promoter regions. Although an NDR was established *in vitro*, the depletion was less pronounced *in vitro* than *in vivo*. Additionally, the strong positioning of the -1 and +1 nucleosomes was not observed (Kaplan *et al.*, 2009; Zhang *et al.*, 2009). One key experiment was performed using human cells and revealed that *in vitro* GC-rich promoters were favored in nucleosome formation, whereas they are depleted *in vivo* (Valouev *et al.*, 2011). Additionally, when the strong Widom601 sequence was inserted into the yeast genome, no strongly positioned nucleosomes over the sequence have been detected (Perales *et al.*, 2011). These results indicate that besides DNA sequence subsidiary factors like ATP-dependent chromatin remodelers, transcription factors and active transcription influence the nucleosome pattern *in vivo*.

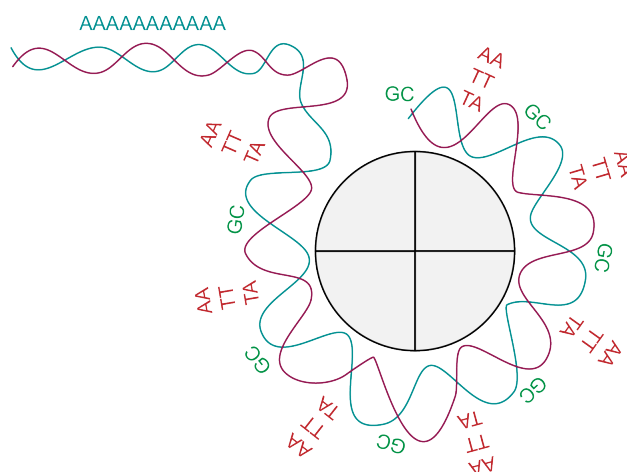


Figure 1.5 Sequence preferences during nucleosome formation.

Illustration of nucleosomal DNA (cyan and pink) wrapped around the histone octamer (grey). Nucleosome formation favors DNA sequences that contain AA/TT/TA-dinucleotides (red) in a 10 bp periodicity in the minor groove facing to the histone octamer and GC-dinucleotides (green) facing outward. Stretches of homopolymeric As and Ts (cyan) are disfavored. Redrawn from (Struhl and Segal, 2013).

1.2.4.2 Chromatin remodelers

When adding yeast cell extract and ATP to purified histones and DNA, the nucleosome positioning pattern resembles the *in vivo* nucleosome positioning pattern more closely (Korber and Hörz, 2004; Zhang *et al.*, 2011): the depletion of nucleosomes at promoters is enhanced, comparable to the *in vivo* level and well-positioned +1 and -1 nucleosomes are generated. This effect is highly ATP-dependent and suggests an involvement of ATP-dependent chromatin remodeling complexes (reviewed in (Clapier and Cairns, 2009; Zhang *et al.*, 2011)). There are four different chromatin remodeling complex families known to date, which are able to engage, select and remodel nucleosomes: SWI/SNF (switching defective/sucrose nonfermenting), ISWI (imitation switch), CHD (chromodomain, helicase, DNA binding), INO80 (inositol requiring 80). Chromatin remodeler complexes have several common characteristics: i) affinity to the nucleosome itself, ii) a histone modification recognition domain, iii) a similar DNA-dependent ATPase domain to break histone-DNA contacts prior to remodeling, iv) ATPase regulating domains and proteins and v) domains and proteins to interact with other factors. However, the individual complexes harbor unique domains, which allow their separation into distinct families (Table 1.3). The individual families are evolutionarily conserved and best studied in human, yeast, fly, mouse, frog and plants, although the individual composition in each organism varies.

Table 1.3 ATP-dependent chromatin remodeler complexes in *S. cerevisiae*.

List of yeast chromatin remodeling complexes and their ATPase assigned to the respective family. Modified from (Clapier and Cairns, 2009). HSA, helicase-SANT; SANT, Swi3, Ada2, N-Cor, and TFIIB; SLIDE, SANT-like ISWI domain.

Family	SWI/SNF		INO80		ISWI			CHD
Complex	SWI/SNF	RSC	SWR1	INO80	ISW1a	ISW1b	ISW2	CHD1
ATPase	Swi2/Snf2	Sth1	Swr1	Ino80	Isw1		Isw2	Chd1
Additional ATPase domains	bromodomain, HSA		HSA		SANT, SLIDE			chromodomain, DNA binding
Reference	(Mohrmann and Verrijzer, 2005; Bao and Shen, 2007)		{Bao and Shen, 2007, #79911}		{Corona and Tamkun, 2004, #46418}			{Marfella and Imbalzano, 2007, #18475}{Mohrmann and Verrijzer, 2005; Bao and Shen, 2007}

In general, chromatin remodelers utilize the released energy resulting from ATP hydrolysis to alter the structure, position or composition of nucleosomes. This is facilitated by the nucleosomal DNA forming a loop, which is translocated by the ATPase, then the histone octamer can be either slid along the DNA or evicted from the nucleosome (reviewed in Saha *et al.*, 2006; Cairns, 2007). Chromatin remodelers are involved in various cellular processes such as DNA replication, DNA repair and especially in transcription, where they shape the nucleosomal landscape around promoters. The mechanistic details are specific for each remodeler complex.

The complexes of the SWI/SNF family, SWI/SNF and RSC, alter the structure of the chromatin by sliding and/or ejecting nucleosomes independently of transcription (Hirschhorn *et al.*, 1992; Venters and Pugh, 2009b). The ATPases of both, SWI/SNF and RSC, contain a bromodomain, which binds to acetylated lysines. Since promoter nucleosomes are hyper-acetylated, both complexes are targeted to promoters. SWI/SNF is exclusively found at RNA pol II promoters, whereas RSC is additionally localized to RNA pol I and RNA pol III promoters (Damelin *et al.*, 2002; Ng *et al.*, 2003b). RSC seems to be required for NDR formation, since NDRs are narrowed and the downstream nucleosome array is shifted upstream upon RSC depletion (Ganguli *et al.*, 2014).

The SWR1 complex alters the composition of nucleosomes by replacing the canonical histone H2A with its variant H2A.Z at promoter nucleosomes (Mizuguchi *et al.*, 2004; Raisner *et al.*, 2005), while the INO80 complex removes H2A.Z from nucleosomes (Papamichos-Chronakis *et al.*, 2011).

Members of the ISWI and CHD family are important during nucleosome array formation downstream of the +1 nucleosome and the linkage of the array to the NDR due to their spacing activity (Clapier and Cairns, 2009). Members of the ISWI family negatively regulate transcription, e.g. ISW2 is recruited by Ume6, a regulator of meiotic genes to generate a repressive chromatin structure (Goldmark *et al.*, 2000; Fazzio *et al.*, 2001). The CHD1 is the least well understood remodeling complex. Upon mutation of the complex nucleosome positioning of only few genes was affected, suggesting that CHD1 is targeted to specific genes or that it might work in parallel to other remodelers (Tran *et al.*, 2000). Through its chromodomains, CHD1 interacts with H3K4me3 found at promoters, although this interaction has been shown only *in vitro* so far (Flanagan *et al.*, 2005; Pray-Grant *et al.*, 2005; Biswas *et al.*, 2007). In a pull-down approach CHD1 was found to be part of the SAGA (Spt-Ada-Gcn5 Acetyltransferase) complex (Pray-Grant *et al.*, 2005), which acetylates histones H3, H4 and H2B at promoters.

However, the exact mechanisms how chromatin remodelers influence nucleosome positioning genome-wide is still unclear. When incubating purified histones and DNA with individual chromatin remodelers, the *in vivo* pattern (restricted to individual genes or less precise positioning in general) could only partly be reconstituted suggesting a requirement of an interplay between the different chromatin remodelers (Wippo *et al.*, 2011). Chromatin remodeler complexes can be recruited in many different ways: i) by the NDR itself acting as a long free DNA stretch, ii) by containing a sequence-specific subunit, i.e. Rsc3 of RSC, iii) directly or indirectly through DNA-binding factors like TFs or general regulating factors, iv) by histone PTMs or v) co-transcriptionally via histone PTMs or the RNA polymerase II C-terminal domain (CTD; Lieleg *et al.*, 2014).

1.2.4.3 Transcription factors and active transcription

Besides DNA sequence and chromatin remodelers, active transcription mediated by the binding of transcription factors (TFs) and the transcription machinery have been shown to influence nucleosome positioning. A subset of TFs, the general regulatory factors, e.g. Reb1 in yeast, is able to invade nucleosomal DNA *in vivo* and thereby decreases nucleosome occupancy via recruitment of chromatin remodelers (Yu and Morse, 1999; Raisner *et al.*, 2005). High-resolution *in vivo* nucleosome positioning maps in the absence of active transcription have shown that the +1 nucleosome and downstream nucleosomes are positioned further downstream relative to their original position (Weiner *et al.*, 2010; Zhang *et al.*, 2011).

1.2.4.4 Suggested integrative models

The Korber group proposes the following integrative model regarding the formation of nucleosome architecture around promoters (Lieleg *et al.*, 2014): The NDR is generated by nucleosome-disfavored DNA sequences, e.g. poly(dA.dT) tracts, and chromatin remodelers, e.g. RSC, and general regulatory factors, e.g. Reb1. The width of the NDR is determined by the positioning of the -1 and +1 nucleosomes mainly by recruited chromatin remodelers, e.g. Isw2. The array of regularly spaced nucleosomes downstream of the +1 nucleosome may occur according to statistical positioning against a barrier, but also requires an active process mediated by remodelers. These may be recruited to the barrier and act on individual nucleosomes or as di-nucleosome clamps that dictate the spacing from the barrier. The latter may explain the gradual decrease in positioning within the nucleosome array, since a sufficient amount of remodelers needs to be recruited through active transcription. The final fine-tuning

is mediated by the rotational positioning determined by intrinsic DNA sequence features, e.g. dinucleotide periodicity.

The groups of Struhl and Segal (Hughes *et al.*, 2012; Struhl and Segal, 2013) suggest an integrative model, which adds more importance to active and elongating transcription during the positioning of the +1 nucleosome. They suggest that the RNA pol II pre-initiation complex fine-tunes the positioning of the +1 nucleosome upon positioning by remodelers.

1.2.5 Chromatin regulates post-transcription-initiation events

Besides its regulatory role during transcription initiation, chromatin plays an important role during transcription elongation and RNA maturation. Upon transcription initiation RNA pol II pauses downstream of the promoter and transcription elongation is hindered by the formation of nucleosomes. The following sections address the role of chromatin during promoter-proximal pausing, the mechanism of histone exchange during transcription elongation and the indirect influence of chromatin on RNA maturation during co-transcriptional splicing.

1.2.5.1 Promoter-proximal RNA pol II pausing

The progression of RNA pol II through the promoter-proximal region after transcription initiation provides an additional regulatory mechanism to transcription and is widespread in metazoans (Muse *et al.*, 2007; Zeitlinger *et al.*, 2007; Core *et al.*, 2008; Lee *et al.*, 2008; Gilchrist *et al.*, 2010; Nechaev *et al.*, 2010; Rahl *et al.*, 2010; Min *et al.*, 2011) and lacking in *S. cerevisiae* (Adelman and Lis, 2012).

Upon assembly of the PIC and local unwinding of the DNA, the RNA pol II initiates RNA synthesis and escapes the promoter by releasing the GTFs and the contact to the promoter. During the first step of transcription elongation through the promoter-proximal region RNA pol II pauses within the first 100 bp (Kephart *et al.*, 1992) due to the association of two factors, DSIF (DRB sensitivity inducing factor; Marshall and Price, 1992) and NELF (negative elongation factor; Wada *et al.*, 1998). In *D. melanogaster*, it has been shown that RNA pol II pausing correlates with low nucleosome occupancy at promoters and that nucleosome occupancy increases upon depletion of the pausing factor NELF (Gilchrist *et al.*, 2008, 2010). These findings led to the assumption, that RNA pol II pausing is involved in establishing a transcriptional permissive structure that might allow faster additional transcription initiation in response to specific cues or binding of additional activators. In addition, RNA pol II pausing

ensures a proper protection and maturation of the nascent RNA, since the phosphorylated Ser5 within the CTD of RNA pol II is bound by the 5' capping enzyme (Ghosh *et al.*, 2011).

The transition from the paused to an elongating RNA pol II is mediated by the phosphorylation of the DSIF-NELF complex by the kinase of the positive transcription elongation factor b (P-TEFb; Marshall and Price, 1992, 1995; Wada *et al.*, 1998). Thereby, NELF dissociates from RNA pol II, which allows RNA pol II to elongate the RNA transcript. P-TEFb also phosphorylates Ser2 in the RNA pol II CTD, which finalizes the transition to elongating RNA pol II (Peterlin and Price, 2006).

1.2.5.2 Transcription elongation

Upon transcription initiation RNA pol II needs to overcome the barrier of nucleosomes within the gene body, which hinders progression of RNA pol II. The access of nucleosomal DNA is facilitated by histone exchange, a process that includes the disruption of histone-DNA contacts prior to removal of histones in a sequential manner (Figure 1.6; reviewed in Venkatesh and Workman, 2015). Both, histone-DNA and histone-histone interactions can be weakened by PTMs on histones or by altering the nucleosome composition by replacing canonical histones with histone variants by ATP-dependent chromatin remodelers (Kobor *et al.*, 2004; Smolle and Workman, 2013).

The Ser5 phosphorylation of the CTD of the activated RNA pol II recruits the PAF complex (polymerase II-associated factor; Hampsey and Reinberg, 2003; Ng *et al.*, 2003c) and the Bur1/2 (bypass UAS requirement) complex. Bur1 phosphorylates the E2 ligase Rad6 (Wood *et al.*, 2005), which mono-ubiquitinates the E3 ligase Bre1. Bre1 transfers the ubiquitin to H2BK123 (Robzyk *et al.*, 2000; Ng *et al.*, 2003b; Wood *et al.*, 2003; Kao *et al.*, 2004; Laribee *et al.*, 2005; Wood *et al.*, 2005), which mediates the tri-methylation of H3K4 by COMPASS/Set1 (Complex of proteins associated with Set1 (Su(var)3-9, Enhancer-of-zeste, Trithorax 1) and H3K79 by Dot1 (disruptor of telomeric silencing 1; Briggs *et al.*, 2002; Dover *et al.*, 2002; Sun and Allis, 2002; Wood *et al.*, 2003; Nakanishi *et al.*, 2008). H3K4me3 is then bound by the chromodomain of the acetyltransferase NuA4, which acetylates the H4K12 and stimulates the recruitment of the bromodomain-containing SAGA complex (Ginsburg *et al.*, 2014), which acetylates H3.

The acetylation of promoter-histones plays a crucial role in regulating the chromatin structure around active promoters. It aids the RSC complex to maintain the NDR and recruits together with the NDR the SWR complex via its bromodomain-containing subunit Bdf1 (Durant and Pugh, 2007), which also recruits the GTF TFIID and thereby facilitates the PIC assembly.

The SWR complex replaces sequentially the H2A-H2B dimers with dimers composed of H2B and the histone variant H2A.Z (Wu *et al.*, 2005).

The histone variant H2A.Z shares a ~60% sequence identity with its canonical counterpart H2A and its sequence is highly conserved among species (Zlatanova and Thakar, 2008; Talbert and Henikoff, 2010). H2A.Z is incorporated into the nucleosome in a replication-independent manner (Mizuguchi *et al.*, 2004). Its incorporation affects the interface between the H2A.Z-H2B dimer and the H3-H4 tetramer, which leads to changes in the biochemical properties in the nucleosome affecting PTMs, protein interactions and chromatin structure (Talbert and Henikoff, 2010) and thereby decreases nucleosome stability (Suto *et al.*, 2000).

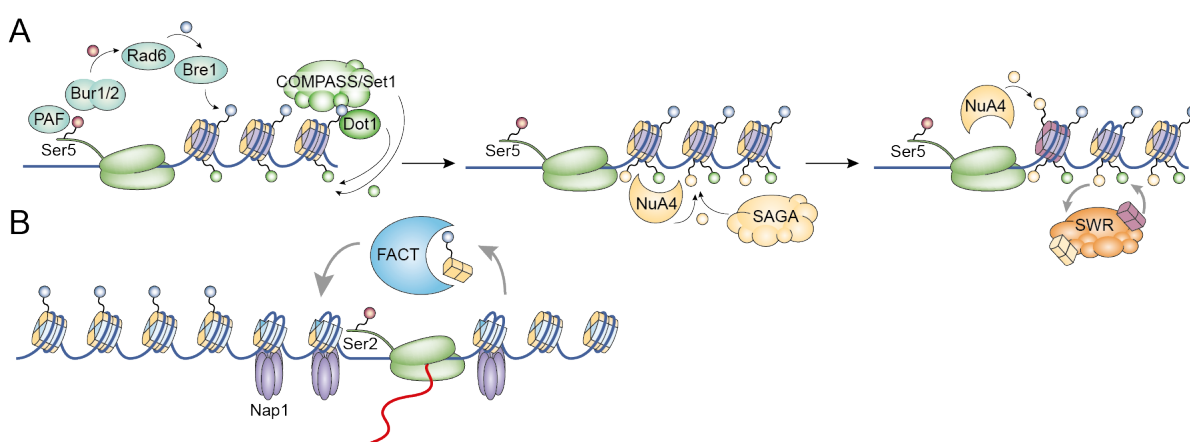


Figure 1.6 Mechanism of histone exchange.

(A) Histone exchange upon transcription initiation. PAF1 and Bur1/2 bind to the phosphorylated Ser5 within the CTD of RNA pol II and Bur1 phosphorylates Rad6. Rad6 mono-ubiquitinates Bre1, which transfers the ubiquitin to H2BK123 stimulating COMPASS/Set1 and Dot1 to tri-methylate H3K4 and H3K79, resp. H3K4me3 is bound by NuA4, which acetylates H4K12 and stimulates the recruitment of the SAGA complex, which acetylates several lysines within the N-terminal of H3. The acetylation of promoter-histones recruits the SWR complex, which replaces sequentially the H2A-H2B dimers with H2A.Z-H2B dimers. In addition, SWR interacts with NuA4, which acetylates H2A.Z on K14, upon incorporation. (B) Histone exchange during transcription elongation. The ubiquitinylation of H2BK123 facilitates the eviction of one H2A-H2B dimer via the FACT histone chaperone complex and Nap1 stabilizes the remaining histone hexamer. Modified from (Venkatesh and Workman, 2015).

The SWR complex interacts with NuA4, which acetylates H2A.Z on K14, once it is incorporated into the promoter nucleosome (Millar *et al.*, 2006). This prevents the removal of H2A.Z from the nucleosome by INO80 ensuring the retention of H2A.Z within promoter nucleosomes (Papamichos-Chronakis *et al.*, 2011). Once transcription has started and RNA pol II has overcome the instable H2A.Z-containing nucleosomes its progression is ensured by the eviction of one H2A-H2B dimer (Kireeva *et al.*, 2002; Belotserkovskaya *et al.*, 2003; Kulaeva *et al.*, 2009). This is facilitated by the ubiquitinylation of H2BK123 and the FACT histone chaperone complex (Pavri *et al.*, 2006). The resulting hexamer composed of the H3-H4 tetramer and the remaining H2A-H2B dimer is stabilized by the histone chaperone Nap1

(Kuryan *et al.*, 2012). Chaperones are proteins that interact with histones and are involved in a broad spectrum of processes, like histone transport, storage and nucleosome assembly and disassembly.

1.2.5.3 Splicing

Variations in the degree of nucleosome occupancy are not only found at promoters, but also at DNA motifs important for RNA splicing. During splicing the nascent and precursor mRNA (pre-mRNA) is processed into mature messenger RNA (mRNA), which after additional modifications is the template for translation to produce proteins. During transcription elongation the 5'-end of the pre-mRNA is stabilized and protected against degradation by the addition of a 7-methylguanosin, the so-called cap. The emerging capped pre-mRNA is composed of alternating intron and exon sequences, whereas only exons contain coding sequences. During splicing intron sequences are co-transcriptionally removed and exon sequences are joint together yielding mRNA (Figure 1.7). Intron sequences are marked at the 5'-end with a conserved GU-dinucleotide (splice donor site) and at the 3'-end with a conserved AG-dinucleotide (splice acceptor site), which is preceded by a pyrimidine-rich polyY tract and an A as branch point. In a first reaction, the splice donor site forms a 2'-5'-phosphodiester bond with the branch point (Guth *et al.*, 2001) forming a lariat-like structure (Ruskin *et al.*, 1984). This step is facilitated amongst other proteins by U1 snRNP (small nuclear ribonucleoprotein) bound to the splice donor site and U2 snRNP bound to the branch point. U2 snRNP requires the previous binding of U2AF, which consists of a U2AF65 subunit (U2 auxiliary factor, binds the polyY tract) and a U2AF35 subunit (binds to the splice acceptor site; Zamore *et al.*, 1992; Wu *et al.*, 1999; Kielkopf *et al.*, 2001). Upon the first reaction, U6 snRNP replaces U1 snRNP at the free splice donor site (Staley and Guthrie, 1998) and forms the active site for the second reaction, in which the splice donor site and the splice acceptor site are ligated to join the exons and the lariat-shaped intron is released (Madhani and Guthrie, 1994).

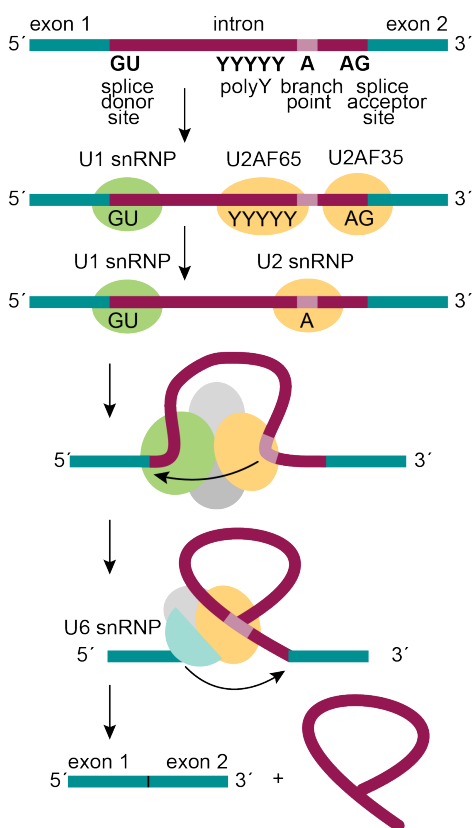


Figure 1.7 Mechanism of RNA splicing.

The coding sequence of pre-mRNA is split among exons (cyan), which are interspaced by non-coding introns (pink). Introns are marked at each end with conserved sequences: a GU at the 5'-end serving as splice donor site (SDS) and an AG at the 3'-end serving as splice acceptor site (SAS). The SAS is preceded by an adenosine serving as branchpoint (highlighted in light pink) and a pyrimidine-rich polyY tract. The SDS, polyY tract and SAS are bound by U1 snRNP, U2AF65 and U2AF35, resp. U2snRNP binds to the branch point and U2AF65 and U2AF35 are released. Upon an interaction between U1 snRNP, U2 snRNP and other proteins, the SDS and the branch point form a 2'-5' phosphodiester bond, which results in a lariat-like structure. The ligation of both exons is mediated by U6 snRNP and releases the lariat-shaped intron.

In genome-wide studies exons have been shown to be highly occupied by nucleosomes compared to introns, which has been linked to the higher GC-content within exons (Schwartz *et al.*, 2009). In addition, the presence of polypyrimidine-rich sequences that are part of splice signals located at intron/exon boundaries have been shown to be nucleosome depleted (Figure 1.8; Schwartz *et al.*, 2009; Tilgner *et al.*, 2009; Chen *et al.*, 2010). Both, high nucleosome occupancy throughout exons and nucleosome depletion at intron/exon boundaries are evolutionarily conserved (Gelfman *et al.*, 2012).

The formation and maintenance of this prominent chromatin structure around splice sites also involves PTMs and the action of chromatin remodelers. In yeast, exons are marked by H3K4me3, which has been shown to recruit the chromatin remodeler CHD1, whose depletion decreases splicing efficiency drastically (Sims *et al.*, 2007). For another chromatin

remodeler, SWI/SNF, it has been shown that its ATPase catalytic subunit Brm interacts with a subunit of the spliceosome that is involved in exon recognition (Batsché *et al.*, 2006).

In the context that splicing occurs co-transcriptionally, a speed bump model has been proposed in which nucleosomes act as barriers slowing transcription elongation. This allows the assembly of the splicing machinery at the slowly emerging pre-mRNA and ensures a proper exon inclusion into the mRNA (Schwartz and Ast, 2010). Thereby, chromatin structure influences splicing efficiency and gene expression. Moreover, it has been suggested that high nucleosome occupancy at exons protects coding sequences from mutational agents (Tolstorukov *et al.*, 2011).

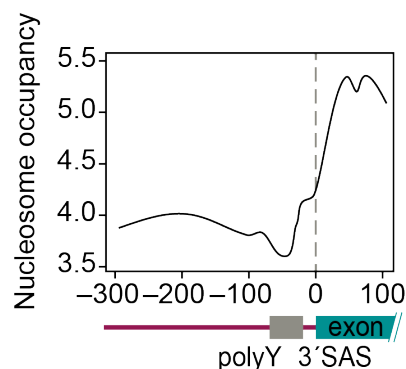


Figure 1.8 Nucleosome occupancy across the 3' splice acceptor site (SAS).

Nucleosome occupancy levels in activated T-cells aligned to the 3' SAS (dashed line). Introns, exons and polyY tracts are shown in pink, cyan and grey. Redrawn from (Schwartz *et al.*, 2009).

1.3 Focused and dispersed transcription initiation

The findings outlined above concern mostly TATA-box-containing promoters which promote transcription initiated from a single TSS, a process referred to as focused transcription. Promoters driving focused transcription, in the following referred to as focused promoters, typically contain well-defined promoter elements and are present in all organisms studied thus far (Juven-Gershon and Kadonaga, 2010). However, with the development of more and more tools to study gene expression on a genome-wide scale, it is becoming increasingly clear that ~70% of mammalian genes are regulated by dispersed promoters (Saxonov *et al.*, 2006). Here, TSSs are spread over 50-100 bp, sometimes even across regions spanning 10 kb (Koch *et al.*, 2011). Dispersed promoters are often found in CpG islands and typically lack well-defined promoter elements (Carninci *et al.*, 2006; Saxonov *et al.*, 2006; Sandelin *et al.*, 2007). Besides mammals, dispersed promoters have been identified in *S. cerevisiae* (Zhang and Dietrich, 2005), *Arabidopsis thaliana* (Yamamoto *et al.*, 2009), *Drosophila melanogaster* (Ni *et al.*,

2010), *Xenopus laevis* (van Heeringen *et al.*, 2011), *Leishmania major* (Martínez-Calvillo *et al.*, 2003) and *Trypanosoma brucei* (Kolev *et al.*, 2010).

Compared to focused promoters, dispersed promoters appear to be more enriched in H2A.Z (Rach *et al.*, 2011). Although their high GC content should provide a favored DNA sequence for nucleosome formation (Ramirez-Carrozzi *et al.*, 2009; Tillo *et al.*, 2010; Valouev *et al.*, 2011) dispersed promoters show a static chromatin structure with broader NDRs at their TSSs (Tirosh and Barkai, 2008) and thereby tend to have a more open chromatin structure (Jones, 2012). This is additionally supported by the enrichment of RNA pol II due to promoter-proximal pausing, which counteracts the tendency of nucleosome formation within promoters (Core *et al.*, 2008). Based on these findings it has been suggested that the activity of dispersed promoters within CpG islands is independent of chromatin remodelers (Ramirez-Carrozzi *et al.*, 2009).

Although it is not known how exactly PICs assemble at CpG islands (Luse, 2014), CpG islands and dispersed promoters in general share some common characteristics with focused promoters during transcription initiation. They are marked by H3K4me3 (Guttman *et al.*, 2009), contain transcription factor binding sites (Landolin *et al.*, 2010) and RNA pol II undergoes promoter-proximal pausing (Hargreaves *et al.*, 2009).

Dispersed promoters tend to be located upstream of constitutively expressed genes, whereas focused promoters are associated upstream of regulated genes (Juven-Gershon and Kadonaga, 2010). This correlation has raised questions about the requirement of defined promoter motifs during transcription initiation of constitutively expressed genes or about the existence of promoter motifs in organisms that do not regulate transcription initiation.

1.4 *Trypanosoma brucei*

1.4.1 General overview

The flagellated protozoan parasite *Trypanosoma brucei* belongs to the class of kinetoplastea and the order of Trypanosomatida and has branched from the eukaryotic lineage early in evolution (Fernandes *et al.*, 1993). *T. brucei* is the causative agent of the vector-borne disease African trypanosomiasis, which affects humans (sleeping sickness) and cattle (nagana). In 2009, ~10,000 human cases have been reported with ~3000 annual infections per year (WHO, 2015). Transmission of the disease is restricted to the habitat of its vector, the tsetse fly (*Glossina* sp.; Malvy and Chappuis, 2011), which occurs almost exclusively in Sub-Saharan Africa. The species *Trypanosoma brucei* encompasses three infective subspecies:

i) *Trypanosoma brucei gambiense*, which causes a chronic course of disease and is found in western and central Africa, ii) *Trypanosoma brucei rhodensiense*, which causes a rapid and acute course of disease and is found in eastern and southern Africa and iii) *Trypanosoma brucei brucei*, which infects animals only and is commonly used in the laboratory.

T. brucei resides within its mammalian host extracellularly. Thus, it needs to constantly evade the host immune system due to its exposed surface. The surface of *T. brucei* is coated with ~10 million copies of an identical protein, the variant surface glycoprotein (VSG; Vickerman, 1969; Cross, 1975). Although its genome codes for several thousand distinct VSG genes (Cross *et al.*, 2014). Only one VSG gene is expressed at any time from a specific transcription unit called expression site (Johnson *et al.*, 1987). The genome of *T. brucei* contains ~15 of those expression sites, which are located in the subtelomeric regions of the chromosomes (De Lange and Borst, 1982; Hertz-Fowler *et al.*, 2008) and only one is active at any time while the remaining expression sites are repressed. *T. brucei* uses a mechanism called antigenic variation to facilitate a periodic change of the VSG on the trypanosome surface, which repeatedly challenges the host immune system. The exact mechanism however, is not known to date.

1.4.2 Gene expression in *T. brucei*

Trypanosoma brucei and its relatives (order Kinetoplastida, family Trypanosomatidae) *Leishmania* and *Trypanosoma cruzi* own mechanisms to regulate gene expression that distinguish themselves from other eukaryotic systems.

1.4.2.1 Genes are organized in polycistronic transcription units

The diploid genome of *T. brucei* is ~35 Mb in size (haploid) and contains ~9,000 genes that are distributed among 11 chromosomes (Melville *et al.*, 2000; Berriman *et al.*, 2005) and are mostly organized in ~200 polycistronic transcription units (PTUs; Figure 1.9A; Siegel *et al.*, 2009). PTUs contain the majority of protein-coding genes and are transcribed by RNA pol II. A special type of PTUs, the expression sites, are transcribed by RNA pol I. Polycistronic transcription has also been observed in *C. elegans*, however, there PTUs only contain ~15% of the protein-coding genes (Blumenthal *et al.*, 2002). Genes within one PTU are transcribed from the same strand, whereas the direction of neighboring PTUs can be oriented in the same and opposite direction. PTUs can be separated by tandem arrays of RNA pol I-transcribed rRNA or RNA pol III-transcribed tRNA genes, or by so-called strand switch regions, where the transcription sense converges or diverges (Cordingley, 1985; Hernández-Rivas *et al.*, 1992).

1.4.2.2 *trans*-splicing

The polycistronically transcribed RNA is processed into mature mRNA by co-transcriptional *trans*-splicing (Figure 1.9B). Here, the primary polycistronic transcript is separated by a splicing reaction in which a capped 39-nt mini-exon is added to the 5′-end of each 5′UTR (Freistadt *et al.*, 1987; Perry *et al.*, 1987; Freistadt *et al.*, 1988; Bangs *et al.*, 1992). This mini-exon is called the spliced leader (SL) and is transcribed independently by RNA pol II from an array located on chromosome 9 (Kooter and Borst, 1984; Gilinger and Bellofatto, 2001). *Trans*-splicing requires, like *cis*-splicing, a polypyrimidine-rich tract (polyY tract, mostly Ts), a GU dinucleotide at the 5′-splice donor site (SDS) and an AG dinucleotide at the 3′-splice acceptor site (SAS). Both processes follow the same mechanism and most of the factors of the spliceosome are conserved in *T. brucei* (reviewed in Liang *et al.*, 2003).

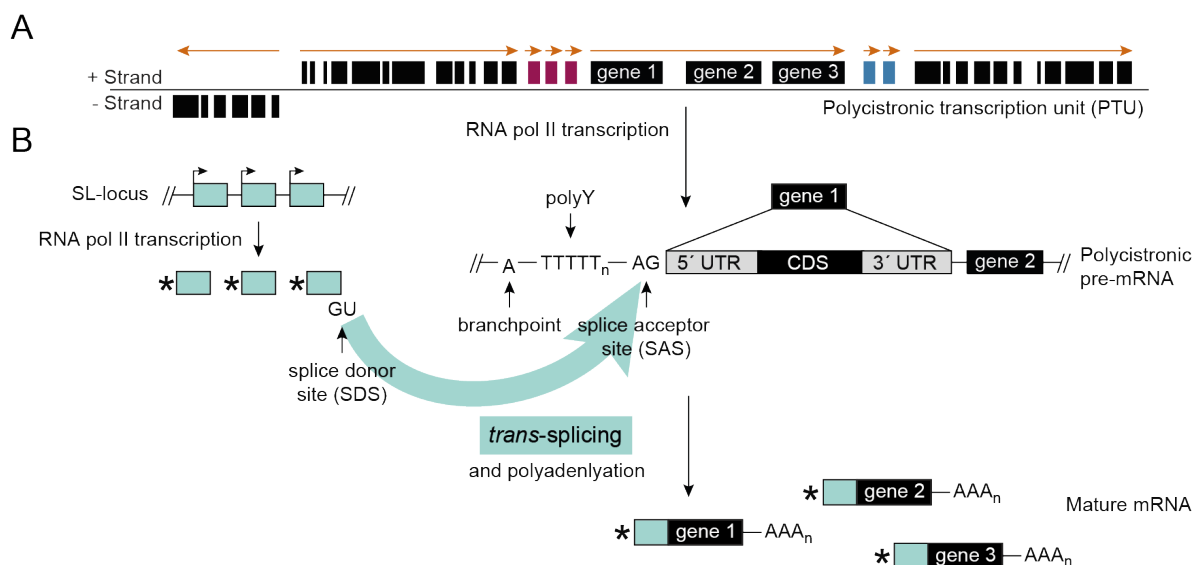


Figure 1.9 Genes are organized in PTUs and *trans*-spliced in *T. brucei*.

(A) Schematic illustration of an exemplary PTU. ORFs, rRNA genes and tRNA genes are shown in black, pink and blue and are transcribed by RNA pol II, RNA pol I and RNA pol III, resp. Orange arrows indicate the direction of transcription. (B) Schematic illustration of pre-mRNA maturation by *trans*-splicing. Genes within the polycistronic pre-mRNA are separated by the splicing of the spliced leader (SL, highlighted in cyan) to the SAS of each gene. The SL is transcribed from a distinct locus and capped (asterisk). After splicing the RNA is polyadenylated and the mature mRNA can be exported from the nucleus.

With the exception of two genes (poly(A) polymerase, DNA/RNA helicase) all genes of the *T. brucei* genome do not contain introns (Mair *et al.*; Berriman *et al.*, 2005). Hence, the 5′-end of each 5′UTR serves as SAS. Systematic mutational analyses, in which the position and composition of polyY tracts and of SASs have been altered, and sequence elements in 5′UTRs have been mutated, revealed the importance of these motifs for efficient *trans*-splicing in

T. brucei (Huang and Van der Ploeg, 1991; Siegel *et al.*, 2005). Since splicing is crucial for the maturation of primary transcripts into mRNA this process provides a mechanism for post-transcriptional regulation of gene expression.

1.4.2.3 Transcription initiation in *T. brucei*

Trypanosoma brucei appears to completely lack the ability to regulate RNA pol II-mediated transcription (Clayton, 2002) and the process of transcription initiation in this organism is still unknown. One of the major reasons might be that, although several attempts have been made, no promoter motifs for RNA pol II-transcribed protein-coding genes have been identified so far. However, based on findings derived from other eukaryotes key players and markers involved in transcription initiation in *T. brucei* have been identified. The following paragraphs summarize the current knowledge about the RNA pol II complex, transcription factors, RNA pol II promoter motifs, the chromatin structure around transcription start sites and chromatin remodelers.

To gain insight whether the mechanism of transcription initiation in *T. brucei* might be similar to the mechanism in *S. cerevisiae*, *in silico* analyses have been performed. In yeast it has been demonstrated that RNA pol I, II and III complexes share 5 common and 7 homologue subunits (Willis, 1993; Geiduschek and Kassavetis, 2001; Hu *et al.*, 2002). The analysis in *T. brucei* revealed that the genome encodes all of the common and most of the homologue subunits among the RNA pol I, II and III complexes (Ivens *et al.*, 2005). Distinct to RNA pol II in most eukaryotes, RNA pol II in *T. brucei* contains a non-canonical CTD that is phosphorylated (Das and Bellofatto, 2009), although it lacks the characteristic heptapeptide repeat that is differentially phosphorylated in the course of transcription (Smith *et al.*, 1989). Regarding the presence of transcription factors, only few have been identified, that are involved in SL RNA transcription. These encompass TRF4 (TBP-related protein 4; Ruan *et al.*, 2004; Das *et al.*, 2005; Schimanski *et al.*, 2005), TFIIB (Palenchar *et al.*, 2006; Schimanski *et al.*, 2006), SNAPc (Das and Bellofatto, 2003; Schimanski *et al.*, 2005), TFIIA (Das *et al.*, 2005; Schimanski *et al.*, 2005) and TFIIF (Lecordier *et al.*, 2007; Lee *et al.*, 2007a).

Amongst all RNA pol II-transcribed genes, only for the SL RNA gene a promoter motif has been identified so far that consists of a bipartite upstream sequence element and an initiator (Gilinger and Bellofatto, 2001). Promoter motifs for RNA pol II-transcribed PTUs are still elusive. Several attempts to identify promoter motifs remained unsuccessful. Putative promoter sequences originating from regions upstream of the RNA pol II-transcribed actin gene cluster (Ben Amar *et al.*, 1991) and *HSP70* (Lee, 1996), resp. have been shown to direct transient expression of a reporter gene. However, it was not clear whether the activity was RNA pol II-mediated and follow-up studies could not validate this observation (McAndrew *et*

al., 1998). In a second study, the insertion of a putative promoter sequence of *Tbpgt* into mammalian cells resulted in transcriptional activity suggesting a similar transcription initiation mechanism in *T. brucei*, however it is difficult to draw conclusions from heterologous systems (Bayele, 2009). In another study, McAndrew and colleagues observed α -amanitin-sensitive transcription of a reporter gene, which is preceded with a T3 polymerase promoter and inserted in a transcriptional silent region in the genome. The toxin α -amanitin is a strong inhibitor of RNA pol II, whereas other polymerases are either insensitive or only moderately influenced (Schultz and Hall, 1976). Thus, the authors suspected that the active T3 polymerase-mediated transcription leads to an open chromatin conformation and that this allowed RNA pol II to access the DNA and to initiate transcription (McAndrew *et al.*, 1998). Given these results and the lack of transcriptional control (Clayton, 2002), it has been suggested that chromatin structure is a central regulator of transcription initiation in *T. brucei*.

As already described in chapter 1.2 the chromatin structure can be altered by the incorporation of histone variants, PTMs and the action of chromatin remodelers. Albeit histones are evolutionary conserved, there are substantial differences among the trypanosome histones and those of higher eukaryotes regarding their highly modified N-terminal tails (Thatcher and Gorovsky, 1994; Mandava *et al.*, 2007). Many PTMs are trypanosome-specific, while well-conserved PTMs are absent (Janzen *et al.*, 2006a; Mandava *et al.*, 2007). One of the best characterized histone PTM in *T. brucei* is H3K76me, the homologue to K3K79 in other organisms, which is mono-, di- or trimethylated by DOT1 in humans and associated with transcribed chromatin (Steger *et al.*, 2008). In *T. brucei*, H3K76 is mono- and di-methylated by DOT1A and trimethylated by DOT1B (Janzen *et al.*, 2006b). Besides the four canonical histones (H2A, H2B, H3, H4) *T. brucei* expresses four histone variants (H2A.Z, H2B.V, H3.V, H4.V; Alford and Horn, 2004; Lowell and Cross, 2004; Lowell *et al.*, 2005). Histone variants differ from their canonical counterparts in their DNA sequence and are incorporated into nucleosomes in a cell cycle and replication independent manner. The H2A variant H2A.Z is highly conserved among eukaryotes (Malik and Henikoff, 2003) and in *T. brucei* it exclusively dimerizes with the H2B variant H2B.V (Lowell *et al.*, 2005). In genome-wide ChIP-seq assays it has been shown that PTUs are marked by nucleosomes containing H2A.Z/H2B.V at the 5'-end and by nucleosomes containing H3.V/H4.V at the 3'-end (Figure 1.10; Siegel *et al.*, 2009). Due to the instability of H2A.Z/H2B.V-containing nucleosomes (Siegel *et al.*, 2009) it has been suggested that those nucleosomes contribute to a more open chromatin structure and thus epigenetically mediate RNA pol II-transcription initiation (Siegel *et al.*, 2009). Compared to other eukaryotes, in which only single nucleosomes around the promoter contain H2A.Z, the sites enriched in H2A.Z/H2B.V-containing nucleosomes span regions of ~10 kb around the 5'-

ends of PTUs in *T. brucei*. Those regions are in the following referred to as ‘transcription start regions’ (TSRs). Besides histone variants (H2A.Z and H2B.V), TSRs are additionally enriched in PTMs (H3K4me3 and H4K10ac) and a bromodomain protein (BDF3) associated with open chromatin (Siegel *et al.*, 2009; Wright *et al.*, 2010). Mapping of primary transcripts genome-wide has revealed that transcription initiates within these regions in a dispersed manner (Kolev *et al.*, 2010). In extensive homology searches within the *T. brucei* genome, several proteins involved in establishing an open chromatin structure could be identified. Those encompass the chromatin remodeler ISWI (de la Serna *et al.*, 2006; Urwyler *et al.*, 2007), histone acetyltransferases (Kawahara *et al.*, 2008; Siegel *et al.*, 2008), several histone methyltransferases (Figueiredo *et al.*, 2009) and several BDFs (Siegel *et al.*, 2009). However, the mechanism, in which these factors act together to establish an open chromatin structure remains still elusive.

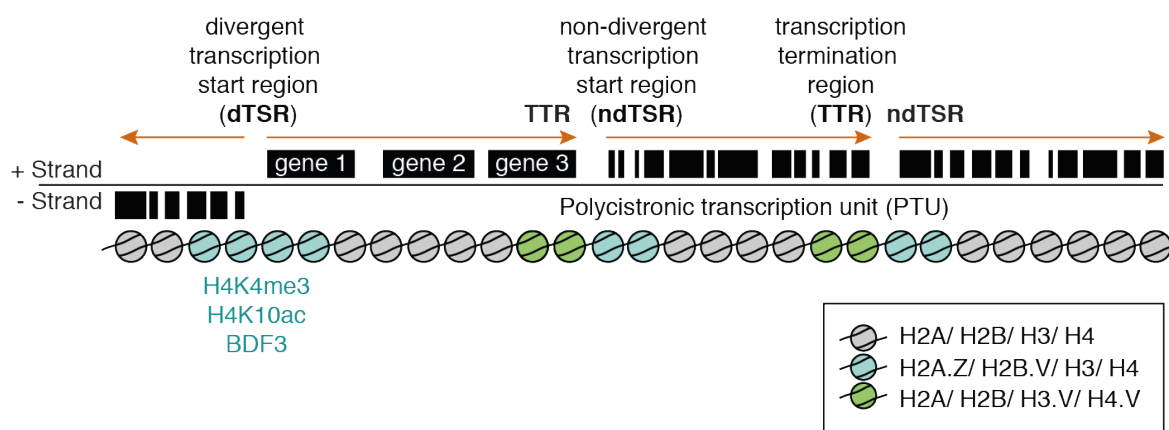


Figure 1.10 Epigenetic marks at PTUs.

Boundaries of PTUs are marked by nucleosomes containing different types of histone variants. Nucleosomes containing H2A.Z and H2B.V (cyan nucleosomes) are enriched at divergent (dTTSRs) and non-divergent transcription start regions (ndTTSRs). Nucleosomes containing H3.V and H4.V (green nucleosomes) are enriched at transcription termination regions (TTRs). Identified TSR-specific PTMs and a chromatin remodeler subunit are shown in cyan. Orange arrows indicate the direction of transcription.

1.5 Aim of the study

In most of the eukaryotic organisms, regulation of gene expression involves transcriptional control. One central regulator during transcription is the DNA sequence as it defines the promoter region necessary for transcription initiation. The nature of regulatory DNA sequences divides promoters into two classes: those containing well defined promoter motifs, e.g. a TATA-box, and those lacking conserved promoter motifs and transcribing constitutively expressed genes. In both classes, however, the DNA sequence provides a binding site for regulatory proteins. Those can either be involved in the establishment of a transcription-permissive

chromatin structure or in the assembly of the RNA polymerase complex and thereby regulate transcription initiation. This raises the question whether defined promoter motifs are required for the expression of constitutively expressed genes and whether defined promoter motifs exist in organisms that do not regulate transcription initiation, such as trypanosomes. Thus, this study aimed to shed light on the regulation of gene expression in *T. brucei* with a focus on the role of the DNA sequence by addressing the following questions:

- Where does RNA pol II transcription initiate within TSRs?
- Does the chromatin structure across TSRs differ to that at other genomic sites?
- Is the DNA sequence found within TSRs sufficient to promote transcription initiation and is it possible to identify DNA sequence motifs?
- Is the DNA sequence sufficient to promote targeted histone variant deposition?
- Does the *T. brucei* genome contain NDRs and if yes, are they involved in regulation of gene expression?
- Does the DNA sequence affect nucleosome positioning?

Answering these questions will provide insights into the previously little investigated mechanism of dispersed transcription initiation and regulation of gene expression in an early diverged eukaryotic organism.

2

Materials and methods

2.1	Molecular cloning methods	45
2.2	Generation of constructs	48
2.3	Trypanosome cell culture and analysis.....	61
2.4	Biochemical methods.....	67
2.5	Next-generation sequencing methods.....	69
2.6	Data generated in this study and source code availability	77
2.7	Software	77

2.1 Molecular cloning methods

2.1.1 Polymerase chain reaction (PCR)

To amplify DNA fragments to be inserted in a plasmid backbone a 50 μ l reaction was set up containing 25 ng of plasmid DNA or 100 ng of genomic DNA from *T. brucei*, 0.2 mM of each dNTP (Thermo Scientific), 0.5 μ M of each specific forward and reverse primer (synthesized by Sigma Aldrich), 1 U of Phusion High-Fidelity DNA Polymerase (Thermo Scientific), Phusion HF Buffer and dH₂O. The cycling conditions were adjusted according to the melting temperature of the primers and the amplicon length: 98 °C/30 sec, 25 cycles (98 °C/10 sec – X °C/30 sec – 72 °C/30 sec/kb), 72 °C/5 min, 12 °C/hold.

To verify bacterial colonies after transformation (see chapter 2.1.3 and 2.1.4) a 20 μ l reaction was set up containing 0.2 mM of each dNTP (Thermo Scientific), 0.5 μ M of each specific forward and reverse primer (synthesized by Sigma Aldrich), 10% DMSO, 1 U of DreamTaq DNA Polymerase (Thermo Scientific), DreamTaq Buffer and dH₂O. Per reaction one colony was picked with a pipette tip and transferred to the PCR tube by rubbing on the tube wall. Remaining bacteria on the tip were streaked on a LB-agar plate supplemented with 50 μ g/ml Ampicillin and incubated over night at 37 °C. The cycling conditions were adjusted according to the melting temperature of the primers and the amplicon length: 95 °C/5 min, 30 cycles (95 °C/3 min – X °C/30 sec – 72 °C/1 min/kb), 72 °C/5 min, 12 °C/hold.

To verify the integration of a transfected construct into the *T. brucei* genome integration PCRs were performed using 1 μ l of genomic DNA (prepared with Phusion Human Specimen Direct PCR Kit, see chapter 2.3.5) and DreamTaq DNA Polymerase as described above. The cycling conditions were adjusted according to the melting temperature of the primers and the amplicon length: 95 °C/3 min, 30 cycles (95 °C/30 sec – X °C/30 sec – 72 °C/1 min/kb), 72 °C/5 min, 12 °C/hold.

2.1.2 Restriction digest

To digest backbones and inserts prior to ligation or InFusion reaction, 2 μ g of DNA were digested in a 50 μ l reaction containing the respective restriction enzyme/s (New England Biolabs), the corresponding buffer and dH₂O. The reaction was incubated for 1 hour at the temperature according to the enzymes' requirements.

To linearize plasmids prior to transfection 50 μ g of DNA were digested in a 100 μ l reaction containing the respective restriction enzyme/s (New England Biolabs), the

corresponding buffer and dH₂O. The reaction was incubated for 3 hours at the temperature according to the enzymes' requirements.

2.1.3 InFusion and transformation

For InFusion reactions, the backbones were digested according to the descriptions in chapter 2.1.2. Inserts were amplified as described in chapter 2.1.1 using oligonucleotides containing overhangs complementary to the first 15 bp of each end of the linearized backbone vector as primers. Backbones and inserts were purified using the NucleoSpin Gel and PCR Clean-up Kit according to the instructions of the manufacturer (Macherey&Nagel). InFusion reactions were performed using the InFusion® HD Cloning Plus reagents according to the manufacturer's instructions (Clontech Laboratories) with minor changes. In brief, 50 ng of backbone, 25 ng of insert and 1 μ l of InFusion HD Enzyme Premix were mixed and the total volume was adjusted to 5 μ l with dH₂O. The reaction was incubated for 15 min at 50 °C and placed on ice. Stellar competent cells were thawed gently on ice, 50 μ l transferred to a 15 ml reaction tube and 2.5 μ l of InFusion reaction were added. After 30 min on ice a heat shock for 45 sec at 42 °C was performed. After 2 min incubation on ice 445 μ l of SOC medium were added and the bacteria were allowed to recover for 1 hour at 37 °C and 200 rpm. 20 μ l of the bacteria were spread on a LB-agar plate supplemented with 50 μ g/ml Ampicillin. The remaining bacteria were centrifuged at 3500 rpm for 1 min, the supernatant removed except of ~100 μ l, the pellet resuspended and spread on a LB-agar plate supplemented with 50 μ g/ml Ampicillin. Colonies were tested via colony PCR (see chapter 2.1.1) or used to set up a liquid overnight culture in LB medium to isolate plasmids (see chapter 2.1.5) and to perform restriction digests (see chapter 2.1.2).

2.1.4 Ligation and transformation

For ligation reactions, the backbones and inserts were digested according to the descriptions in chapter 2.1.2 and purified using the NucleoSpin Gel and PCR Clean-up Kit according to the instructions of the manufacturer (Macherey&Nagel). Ligation reactions were performed using the T4 DNA Ligase according to the manufacturer's instructions (Thermo Scientific). In brief, insert and 100 ng of backbone were mixed in a molar ratio to of 5:1, T4 DNA Ligase Buffer and 1 U of T4 DNA Ligase were added and the total volume was adjusted to 20 μ l with dH₂O. The reaction was incubated for 1.5 h at RT. Top10 competent cells were thawed gently on ice, 30 μ l transferred in a 1.5 ml reaction tube and 6 μ l of the ligation reaction were added. After

30 min on ice a heat shock for 1 min at 42 °C was performed. After 1 min incubation on ice 300 μ l of SOC medium was added and the bacteria were allowed to recover for 1 hour at 37 °C and 200 rpm. 20 μ l of the bacteria were spread on a LB-agar plate supplemented with 50 μ g/ml Ampicillin. The remaining bacteria were centrifuged at 3500 rpm for 1 min, the supernatant removed except of \sim 100 μ l, the pellet resuspended and spread on a LB-agar plate supplemented with 50 μ g/ml Ampicillin, which have been incubated over night at 37 °C. Colonies were tested via colony PCR (see chapter 2.1.1) or used to set up a liquid overnight culture in LB medium to isolate plasmids (see chapter 2.1.5) and to perform restriction digests (see chapter 2.1.2) or submit for Sanger sequencing (see chapter 2.1.6).

2.1.5 Plasmid isolation

Plasmids were isolated from 4 ml of liquid bacterial overnight culture using the NucleoSpin Plasmid Kit or from 100 ml of liquid bacterial overnight culture using the NucleoBond Xtra Midi Kit from Macherey&Nagel according to the instructions of the manufacturer.

2.1.6 Sanger sequencing

To verify generated constructs via Sanger sequencing 500 ng of plasmid were mixed with 2.5 μ M of specific primer in a total volume of 10 μ l and send to the sequencing service Macrogen.

2.1.7 Bacterial stock preparation

Bacterial strains were preserved in 20% glycerol by adding 300 μ l of 50% glycerol to 700 μ l of a liquid bacterial overnight culture and immediate freezing at -80 °C.

2.1.8 EtOH precipitation

To concentrate and sterilize digested constructs prior to transfection, the DNA was first purified using the NucleoSpin Gel and PCR Clean-up Kit (Macherey&Nagel) with minor changes: each sample was split among 3 columns and eluted twice with 34 μ l Buffer NE, each. Then, the DNA was precipitated as follows: 1/10 of the total volume of 3 M NaOAc and 2.5 V of 100% EtOH were added and the DNA was incubated for 30 min on ice. The precipitated DNA was

pelleted by centrifugation at 16,000 xg for 30 min, washed with 70% EtOH and dried under sterile conditions. The DNA was resuspended in 20 μ l of sterile dH₂O.

2.2 Generation of constructs

All parental and generated constructs used in this study are listed in Table 2.1 and Table 2.2. If not mentioned otherwise all cloning reactions have been performed using InFusion HD Cloning Plus reagents (see chapter 2.1.3). Oligonucleotides were synthesized by Sigma Aldrich and are listed in Table 2.3. gBlocks were synthesized by Integrated DNA Technologies (IDT) and are listed in Table 2.4.

Table 2.1 List of parental constructs used in this study.

Abbreviations: Amp, ampicillin; BLE, phleomycin resistance gene; PAC, puromycin N-acetyltransferase; Tet, tetracycline; Hygro, hygromycin; Blas, blasticidin; Phleo, phleomycin; Puro, puromycin.

Name	Purpose	Resistance in <i>E. coli</i>	Resistance in <i>T. brucei</i>	Reference
pPOTv3_TY-H3V	Retrieve backbone for pPOTv3_TY-RPB9_BSD	Amp	Blas	unpublished, AJ Kraus
pLEW100v5	Amplification of <i>BLE</i>	Amp	Phleo	George Cross, Addgene #24012
pyrFEKO-PUR	Endogenous knock out and flanking of integration site with loxP sites, amplification of <i>PAC</i>	Amp	Puro	(Scahill <i>et al.</i> , 2008)
pLEW111	Tet-inducible ectopic overexpression of target gene from rRNA locus	Amp	Phleo	(Hoek <i>et al.</i> , 2000)
pyrFEKO-HYG	Endogenous knock out and flanking of integration site with loxP sites	Amp	Hygro	(Scahill <i>et al.</i> , 2008)
pLEW100v5_HYG	Tet-inducible ectopic overexpression of target gene from rRNA locus	Amp	Hygro	George Cross, Addgene #24012
pCJ25ARluc	Expression of Rluc from the VSG pseudogene in the active 221 BES	Amp	Blas	unpublished, CJ Janzen
pLEW100Cre-EP1	Tet-inducible Cre-recombinase expression	Amp	Phleo	(Scahill <i>et al.</i> , 2008)

2.2.1 Generation of pPOTv3_TY-RPB9_Phleo/Puro

The constructs pPOTv3_TY-RPB9_Phleo and pPOTv3_TY-RPB9_Puro were generated to N-terminally tag both endogenous alleles of the RNA pol II complex subunit RPB9 (TriTrypDB GeneID Tb427tmp.02.5180) with two Ty1 epitope tags using the pPOTv3 system (Dean *et al.*, 2015). To insert the upstream homology region, 300 bp upstream of the start codon of RPB9 were amplified using oCW_1 and oCW_2 and inserted into pPOTv3_TY-H3V upon digestion with Apal and NotI. After transformation the colonies were verified via colony PCR using the

oligonucleotides oCW_1 and oCW_3. To insert the downstream homology region, 300 bp downstream of the start codon of RPB9 were amplified using oCW_4 and oCW_5 and inserted into the product of the previous cloning step upon digestion with *SacI* and *NheI* generating pPOTv3_TY-RPB9_BSD. After transformation the colonies were verified via colony PCR using the oligonucleotides oCW_4 and oCW_6. Depending on the genetic background of the acceptor cell line the resistance gene was exchanged by a phleomycin resistance gene (amplified with oCW_7 and oCW_8 from pLEW100v5) and a puromycin resistance gene (amplified with oCW_9 and oCW_10 from pyrFEKO-PUR) after pPOTv3_TY-RPB9_BSD was digested with *NotI* and *MluI* generating pPOTv3_TY-RPB9_Phleo and pPOTv3_TY-RPB9_Puro, respectively. After transformation the colonies were verified via colony PCR using the oligonucleotide pairs oCW_1/oCW_11 and pCW_1/oCW_10, respectively. Both constructs were linearized with *Apal* and *NheI* prior to transfection.

2.2.2 Generation of pLEW111_TY1-H2A.Z

The construct pLEW111_TY1-H2A.Z was generated to constitutively overexpress Ty1-H2A.Z from an rRNA locus. First, the tetracycline operator was removed from pLEW111 (Hoek *et al.*, 2000) by digestion with *BglIII* generating pLEW111-TetOp. The H2A.Z CDS (Tb427.07.6360) was amplified using the oligonucleotides oCW_12 and oCW_13. Upon digestion of both, the PCR product and pLEW111-TetOp with *HindIII* and *BamHI* both fragments were ligated generating pLEW111_TY1-H2A.Z. After transformation the colonies were verified by colony PCR using the oligonucleotides oCW_14 and oCW_13. The construct was linearized with *NotI* prior to transfection.

2.2.3 Generation of pyrFEKO-HYG/PUR_H2A.Z

The constructs pyrFEKO-HYG_H2A.Z and pyrFEKO-PUR_H2A.Z were generated to knock out both endogenous H2A.Z alleles using the pyrFEKO system (Scahill *et al.*, 2008). The downstream region of the H2A.Z CDS was amplified using oCW_15 and oCW_16 and both, the PCR product and pyrFEKO-HYG/PUR were digested with *SbfI* and *BamHI* and ligated. The upstream region of the H2A.Z CDS was amplified using oCW_17 and oCW_18 and both, the PCR product and the product of the previous ligation reaction were digested with *PvuII* and *HindIII* and ligated generating pyrFEKO-HYG_H2A.Z and pyrFEKO-PUR_H2A.Z, respectively. All cloning steps were verified via digest with *SbfI/BamHI* and *PvuII/HindIII*, respectively. Both constructs were linearized with *PvuII* and *SbfI* prior to transfection.

2.2.4 Generation of TSR translocation constructs

The targeting construct pCW24v2 originates from pLEW100v5_HYG (kind gift from George Cross, Addgene plasmid #24012). To generate pCW24v2, the rRNA spacer targeting sequence of pLEW100v5_HYG was removed by digestion with AlwNI and AflII and replaced with a linker sequence (amplified with oCW_19 and oCW_20 from the parental construct) and the upstream homology region Tb427_01_v4:282931-283210 (amplified with oCW_21 and oCW_22 from gDNA). The insertion was verified via colony PCR using the oligos oCW_21 and oCW_23. To insert the downstream homology region Tb427_01_v4:283239-283591 (amplified with oCW_24 and oCW_25 from gDNA), the plasmid was digested with NheI. The insertion was verified via colony PCR using the oligos oCW_23 and oCW_26. Prior to transfection, the plasmid was linearized with NotI.

To generate the no-promoter control (pCW24v2-p), pCW24v2 was digested with BglII. The T7 promoter, which transcribes the selection marker, was reinserted with the annealing product of oCW_27 and oCW_28. The insertion was verified via colony PCR using the oligos oCW_3 and oCW_29. The constructs containing TSR DNA sequences were generated by replacing the rRNA promoter in pCW24v2 by BglII digestion and insertion of fragments amplified using different sets of oligonucleotides (oCW_30-61). The insertion was verified via colony PCR using the oCW_3 and the respective reverse oligo used to amplify the insert.

To generate the targeting construct pCW27v2, the upstream and downstream regions of homology of pCW24v2 were removed with XhoI/AflII and NheI, respectively and replaced by the new upstream homology region Tb427_09_v4:1,067,215-1,067,647 (amplified from gDNA with oCW_62 and oCW_63) and the new downstream homology region Tb427_09_v4:1,067,679-1,068,160 (amplified from gDNA with oCW_64 and oCW_65). The insertion was verified via colony PCR using the oligo pairs oCW_62/oCW_23 and oCW_64/oCW_26. Prior to transfection, the plasmid was linearized with NotI and XhoI.

To generate the targeting construct pCW28v2, the upstream and downstream regions of homology of pCW24v2 were removed with XhoI/AflII and NheI, respectively and replaced by the new upstream homology region Tb427_10_v5:1,926,616-1,927,048 (amplified from gDNA with oCW_66 and oCW_67) and the new downstream homology region Tb427_10_v5:1,927,082-1,927,505 (amplified from gDNA with oCW_68 and oCW_69). The insertion was verified via colony PCR using the oligo pairs oCW_66/oCW_23 and oCW_68/oCW_26. Prior to transfection, the plasmid was linearized with NotI and XhoI.

The respective no-promoter controls (pCW27v2-p, pCW28v2-p), constructs containing TSR DNA sequences and GT-rich promoters have been generated just as described for pCW24v2.

In order to introduce the targeting constructs pCW27v2 and pCW28v2 and their derivatives into *ΔH3.V* cells the resistance marker was exchanged and the constructs were labelled with 'v3'. The hygromycin resistance gene was removed with MscI and SpeI and replaced by a phleomycin resistance gene amplified with oCW_70 and oCW_71.

2.2.5 Generation GT-rich promoter constructs

To generate pCW24v2_GT_210_nt, pCW24v2 was digested with BglII and the synthesized gBlock GT_210_nt containing an Ascl and an FseI restriction site at the 3'-end was introduced. The insertion was verified via colony PCR using the oligos oCW_3 and oCW_72. Digestion with Ascl and FseI and insertion of the synthesized gBlock GT_206_nt fused both synthesized sequence elements creating pCW24v2_GT_416_nt. The construct was verified via colony PCR using the oligos oCW_3 and oCW_73.

To generate pCW24v2_GT_210_nt_rc, the reverse complement sequence of GT_210_nt was amplified with oCW_74 and oCW_75 and inserted into pCW24v2_GT_210_nt upon BglII and Ascl digestion. The construct was verified via colony PCR using the oligos oCW_3 and oCW_75. To generate pCW24v2_GT_416_nt_rc, the reverse complement sequence of GT_416_nt was amplified with oCW_76 and oCW_77 and inserted into pCW24v2_GT_416_nt after BglII and FseI digestion. The construct was verified via colony PCR using the oligos oCW_3 and oCW_77.

To generate pCW24v4 (rRNA promoter control) and pCW24v4-p (no promoter control), pCW24v2 was digested with KpnI and SmaI and the synthesized rRNA_promoter sequence and no-promoter sequence were introduced. Both, the rRNA promoter sequence and no-promoter sequence are preceded by two tetracyclin operators. pCW24v4 and pCW24v4-p were verified via colony PCR using the oligo pairs oCW_3/oCW_78 and oCW_3/oCW_29, respectively.

To generate pCW24v4_GT_210_nt, pCW24v4 was digested with BglII and SmaI and the synthesized GT_210_nt sequence containing an Ascl and an FseI restriction site at the 3'-end was introduced. The insertion was verified via colony PCR using the oligos oCW_3 and oCW_74. Digestion of the plasmid with Ascl and FseI followed by insertion of the synthesized GT_206_nt sequence joined both synthesized sequence elements to create the plasmid

pCW24v4_GT_416_nt. The construct was verified via colony PCR using the oligos oCW_3 and oCW_79.

To generate pCW24v4_GT_210_nt_rc, the reverse complement sequence of GT_210_nt was amplified with oCW_80 and oCW_81 and inserted into pCW24v4_GT_210_nt after BglII and Ascl digestion. The construct was verified via colony PCR using the oligos oCW_3 and oCW_75. To generate pCW24v4_GT_416_nt_rc, the reverse complement sequence of GT_416_nt was amplified with oCW_82 and oCW_83 and inserted into pCW24v4_GT_210_nt after BglII and Ascl digestion. The construct was verified via colony PCR using the oligos oCW_3 and oCW_77.

2.2.6 Generation of pCW37

The construct pCW37 was generated to target *FLUC* to a locus within a PTU on chromosome 1 (Tb427_01_v4:500,640-501,239) to measure endogenous RNA pol II levels. In pCW24v2-p the hygromycin resistance gene was inverted by digestion with SmaI and BglII and insertion of the amplification product of oCW_84 and oCW_85 on pCW24v2-p. The insertion was verified via colony PCR using the oligos oCW_84 and oCW_29. The upstream homology region (Tb427_01_v4:500,640-500,939, amplified with oCW_86 and oCW_87) was exchanged by digestion with XhoI and AflIII and the downstream homology region (Tb427_01_v4:500,940-501,239, amplified with oCW_88 and oCW_89) by digestion with NheI and NotI. The construct was verified via colony PCR using the oligo pairs oCW_86/oCW_23 and oCW_88/oCW_26. Prior to transfection, the plasmid was linearized with NotI and Xho.

2.2.7 Generation of polyY constructs

To study the role of the composition of the polyY tract in gene expression and nucleosome positioning the constructs pCW24v2_optPolyY (contains a long T-rich polyY tract) and pCW24v2_noPolyY (contains no polyY tract) were generated. To this end, pCW24v2_GT_210_nt was digested with SmaI and HindIII to remove the endogenous polyY tract of the GPEET 5'UTR. To insert the respective polyY tract the oligo pairs oCW_90/oCW_91 and oCW_92/oCW_93 were annealed and inserted into the digested backbone. Both constructs were verified via Sanger sequencing using oligo oCW_29 (data not shown) and linearized with NotI prior to transfection.

Table 2.2 List of constructs generated in this study.

Abbreviations: Amp, ampicillin; Blas, blasticidin; Phleo, phleomycin; Puro, puromycin; Hygro, hygromycin.

Construct name	Purpose	Resistance in <i>E. coli</i>	Resistance in <i>T. brucei</i>	Reference
pPOTv3_TY-RPB9_BSD	Endogenous N-terminal tagging of RPB9 with Ty1 epitope tag	Amp	Blas	C Wedel, LSM Müller, R Derr
pPOTv3_TY-RPB9_Phleo	Endogenous N-terminal tagging of RPB9 with Ty1 epitope tag	Amp	Phleo	C Wedel, LSM Müller, R Derr
pPOTv3_TY-RPB9_Puro	Endogenous N-terminal tagging of RPB9 with Ty1 epitope tag	Amp	Puro	C Wedel, LSM Müller, R Derr
pLEW111-TetOp	Constitutive ectopic overexpression of target gene from rRNA locus	Amp	Phleo	J Thürich
pLEW111_TY1-H2AZ	Constitutive ectopic overexpression of Ty1-H2A.Z from rRNA locus	Amp	Phleo	J Thürich
pyrFEKO-HYG_H2A.Z	Endogenous knock out of H2A.Z and flanking of integration site with loxP sites	Amp	Hygro	C Wedel
pyrFEKO-PUR_H2A.Z	Endogenous knock out of H2A.Z and flanking of integration site with loxP sites	Amp	Puro	C Wedel
pCW24v2	Targeting construct to insert rRNA promoter and <i>FLUC</i> between 2 divergent TSRs on chromosome 1 (Tb427_01_v4:283211)	Amp	Hygro	C Wedel, R Derr
pCW24v2-p	Targeting construct without promoter to insert <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regA	Inserts regA (Tb427_10_v5: 800,949-810,167) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regA1	Inserts regA1 (Tb427_10_v5: 800,949-802,743) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regA2	Inserts regA2 (Tb427_10_v5: 802,251-804,045) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regA3	Inserts regA3 (Tb427_10_v5: 803,552-805,345) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regA4	Inserts regA4 (Tb427_10_v5: 804,859-806,648) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regA5	Inserts regA5 (Tb427_10_v5: 806,150-807,942) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regA6	Inserts regA6 (Tb427_10_v5: 807,455-809,249) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regA7	Inserts regA7 (Tb427_10_v5: 808,751-810,167) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regA2rc	Inserts the reverse complement of regA2 (Tb427_10_v5: 802,251-804,045) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr

2. Materials and methods

Construct name	Purpose	Resistance in <i>E. coli</i>	Resistance in <i>T. brucei</i>	Reference
pCW24v2_regB	Inserts regB (Tb427_10_v5: 1,634,960-1,641,653) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regB1	Inserts regB1 (Tb427_10_v5: 1,634,960-1,636,762) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regB2	Inserts regB2 (Tb427_10_v5: 1,636,267-1,638,057) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regB3	Inserts regB3 (Tb427_10_v5: 1,637,571-1,639,366) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regB4	Inserts regB4 (Tb427_10_v5: 1,638,867-1,640,655) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regB5	Inserts regB5 (Tb427_10_v5: 1,640,194-1,641,653) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_regB1rc	Inserts the reverse complement of regB1 (Tb427_10_v5: 1,634,960-1,636,762) and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel, R Derr
pCW24v2_GT_210_nt	Inserts the synthetic promoter sequence GT_210_nt and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel
pCW24v2_GT_210_nt_rc	Inserts the reverse complement of the synthetic promoter sequence GT_210_nt and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel
pCW24v2_GT_416_nt	Inserts the synthetic promoter sequence GT_416_nt and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel
pCW24v2_GT_416_nt_rc	Inserts the reverse complement of the synthetic promoter sequence GT_416_nt and <i>FLUC</i> between 2 divergent TSRs on chromosome 1	Amp	Hygro	C Wedel
pCW27v2	Targeting construct to insert rRNA promoter and <i>FLUC</i> upstream of TSR on chromosome 9 (Tb427_09_v4:1,067,648)	Amp	Hygro	C Wedel
pCW27v2-p	Targeting construct without promoter to insert <i>FLUC</i> upstream of TSR on chromosome 9 (Tb427_09_v4:1,067,648)	Amp	Hygro	C Wedel
pCW27v2_regA	Inserts regA (Tb427_10_v5: 800,949-810,167) and <i>FLUC</i> upstream of TSR on chromosome 9	Amp	Hygro	C Wedel
pCW27v2_regB	Inserts regB (Tb427_10_v5: 1,634,960-1,641,653) and <i>FLUC</i> upstream of TSR on chromosome 9	Amp	Hygro	C Wedel
pCW28v2	Targeting construct to insert rRNA promoter and <i>FLUC</i> within RNA pol III transcribed locus on chromosome 10 (Tb427_10_v5: 1,927,049)	Amp	Hygro	C Wedel
pCW28v2-p	Targeting construct without promoter to insert <i>FLUC</i> within RNA pol III transcribed locus on chromosome 10 (Tb427_10_v5: 1,927,049)	Amp	Hygro	C Wedel
pCW28v2_regA	Inserts regA (Tb427_10_v5: 800,949-810,167) and <i>FLUC</i> within RNA pol III transcribed locus on chromosome 10	Amp	Hygro	C Wedel

2. Materials and methods

Construct name	Purpose	Resistance in <i>E. coli</i>	Resistance in <i>T. brucei</i>	Reference
pCW28v2_regB	Inserts regB (Tb427_10_v5: 1,634,960-1,641,653) and <i>FLUC</i> within RNA pol III transcribed locus on chromosome 10	Amp	Hygro	C Wedel
pCW27v3	To insert pCW27v2 in <i>ΔH3.V</i>	Amp	Phleo	C Wedel
pCW27v3-p	To insert pCW27v2-p in <i>ΔH3.V</i>	Amp	Phleo	C Wedel
pCW27v3_GT_210_nt	To insert pCW27v2_GT_210_nt in <i>ΔH3.V</i>	Amp	Phleo	C Wedel
pCW28v3	To insert pCW28v2 in <i>ΔH3.V</i>	Amp	Phleo	C Wedel
pCW28v3-p	To insert pCW28v2-p in <i>ΔH3.V</i>	Amp	Phleo	C Wedel
pCW28v3_GT_210_nt	To insert pCW28v2_GT_210_nt in <i>ΔH3.V</i>	Amp	Phleo	C Wedel
pCW24v4	pCW24v2, rRNA promoter preceded with 2x TetO	Amp	Hygro	C Wedel
pCW24v4-p	pCW24v2-p, rRNA promoter preceded with 2x TetO	Amp	Hygro	C Wedel
pCW24v4_GT_210_nt	pCW24v2_GT_210_nt, rRNA promoter preceded with 2x TetO	Amp	Hygro	C Wedel
pCW24v4_GT_416_nt	pCW24v2_GT_416_nt, rRNA promoter preceded with 2x TetO	Amp	Hygro	C Wedel
pCW24v4_GT_210_nt_rc	pCW24v2_GT_210_nt_rc, rRNA promoter preceded with 2x TetO	Amp	Hygro	C Wedel
pCW24v4_GT_416_nt_rc	pCW24v2_GT_416_nt_rc, rRNA promoter preceded with 2x TetO	Amp	Hygro	C Wedel
pCW37	Insert <i>FLUC</i> within a PTU to measure endogenous RNA pol II levels	Amp	Hygro	C Wedel
pCW24v2_longPolyY	pCW24v2_GT_210_nt with long T-rich polyY tract instead of endogenous polyY	Amp	Hygro	C Wedel
pCW24v2_noPolyY	pCW24v2_GT_210_nt without polyY tract	Amp	Hygro	C Wedel

Table 2.3 List of oligos used in this study.

Underlined sequences indicate InFusion overhangs; abbreviations: UHR, upstream homology region; DHR, downstream homology region.

Oligo name	Lab internal name	Purpose	DNA sequence 5´-3´
oCW_1	1729_pPOTv3_TY-RPB9_UR_F	RPB9 UHR	<u>CTATAGGGCGAATTGGGCCCTGTGTCAACCAACGGCGATA</u>
oCW_2	1730_pPOTv3_TY-RPB9_UR_R	RPB9 UHR	<u>GCATTATACGCGGCCGCGGTGGTTTTATTCCACCTTAGCTTCG</u>
oCW_3	865_ald3´UTR_R	pPOTv3_TY-RPB9_Phleo verification	CCTCCCCCATCTCCCCTCGAGGCGGAGACTGCAATGCA

2. Materials and methods

Oligo name	Lab internal name	Purpose	DNA sequence 5'-3'
oCW_4	1731_pPOTv3_TY-RPB9_DR_F	RPB9 DHR	<u>CAGGATCGGGTAGTGAGCTCGAGTCCACTTTGACGCAG</u>
oCW_5	1732_pPOTv3_TY-RPB9_DR_R	RPB9 DHR	<u>GCTTTTTCATGCTAGCGTTCACGAAGCATGTGACTTC</u>
oCW_6	404_H2A.Z_dKOP_F	pPOTv3_TY-RPB9_Phleo verification	GAGGCGGTTCGTATCACTACC
oCW_7	1749_pPOTv3_Phleo_F	Phleomycin	<u>AAAACCACCGCGGCCCGCCCCGTACCGGGG</u>
oCW_8	1750_pPOTv3_Phleo_R	Phleomycin	<u>AGCTTGAGCAACGCGTAATACTGCATAGATAACAAACG</u>
oCW_9	1733_pPOTv3_Puro_F	Puromycin	<u>AAAACCACCGCGGCCCGCGTTTCTTACATATTTCTCT</u>
oCW_10	1734_pPOTv3_Puro_R	Puromycin	<u>AGCTTGAGCAACGCGTGGGCTCGAATCCCC</u>
oCW_11	405_H2A.Z_dKOP_R	pPOTv3_TY-RPB9_Puro verification	GCGCGTGAGGAAGAGTTCTT
oCW_12	NS_115_TY1-H2A.Z F	<i>Ty1-H2A.Z</i>	ATATAAGCTTATGGAAGTCCATACTAACCAGGACCCACTTGACTCTCTTACAGGTGATGATGCATG
oCW_13	NS_116_TY1-H2A.Z R	<i>Ty1-H2A.Z</i>	TCTCGGATCCCTACGAGCCCCTCTTCGATTTCT
oCW_14	234_plew111_F	pLEW111-TY1-H2A.Z verification	GAGTGAATCAGGTTAGGG
oCW_15	402_H2A.Z_DR_F	$\Delta H2A.Z$ DHR	TTGTTGCCTTCAGCTCGCTA
oCW_16	403_H2A.Z_DR_R	$\Delta H2A.Z$ DHR	CACTAAAACGGGCCACCTCT
oCW_17	400_H2A.Z_UR_F	$\Delta H2A.Z$ UHR	CGGTACCAACACTAGACGGC
oCW_18	401_H2A.Z_UR_R	$\Delta H2A.Z$ UHR	CGTGTCCGTGTATAATGCGC
oCW_19	1278_pCW24_frgmt1_F	pCW24v2 linker region	<u>TAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTC</u>
oCW_20	1279_pCW24_frgmt1_R	pCW24v2 linker region	CTCGAGCTGCATTAATGAATCGGCCAAC
oCW_21	1503_pCW24_v2_UR_F	pCW24v2 UHR	<u>TTAATGCAGCTCGAGGCGGCCGCTACTCACCTGAGAAGCGGC</u>
oCW_22	1504_pCW24_v2_UR_R	pCW24v2 UHR	<u>ICCGGATAGGCTTAAGCATCGTTGGGTGGCAAAGTG</u>
oCW_23	1508_pCW24_10_1_R	construct verification	GTATTCTGCAAATTTAAATGCTGCTAACA
oCW_24	1505_pCW24_v2_DR_F	pCW24v2 DHR	<u>AGGCATGCAAGCTAGCCGAAACCAAGGCGGAAGAAA</u>
oCW_25	1506_pCW24_v2_DR_R	pCW24v2 DHR	<u>AGAGGATCTGGCTAGGCCGCGCGCATGCGTAGGACTCGGTAT</u>
oCW_26	748_pET_seq_F	construct verification	AGCAGCCCAGTAGTAGGTTG
oCW_27	1407_pCW26_T7prom_plus	T7 promoter	<u>CCGGTACGGGAGATCTCCCTATAGTGAGTTCGTTAATCAAGATCTCCCTATCAGT</u>
oCW_28	1408_pCW26_T7prom_minus	T7 promoter	<u>ACTGATAGGGAGATCTTGATTAATACGACTCACTATAGGGAGATCTCCCGTACCGG</u>

2. Materials and methods

Oligo name	Lab internal name	Purpose	DNA sequence 5'-3'
oCW_29	NS54R_TTSluc	sequence polyY tract	TCTTTATGTTTTTGGCGTCTTCCAT
oCW_30	1546_TSS10_part1_F	regA	<u>CCGGTACGGGAGAT</u> CTCCCTATAGTGAGTCGTATTAATCAAGTTTAGCCCGTTTACCTCC A
oCW_31	1547_TSS10_part1_R	regA	ACTGATAGGGAGATCTAGCTCAGCAGTAATAAGGGTCA
oCW_32	1548_TSS10_part2_F	regA1	<u>CCGGTACGGGAGAT</u> CTCCCTATAGTGAGTCGTATTAATCAAGTTTAGCCCGTTTACCTCC A
oCW_33	1549_TSS10_part2_R	regA1	ACTGATAGGGAGATCTGCGGACACGGATTAGCTGAA
oCW_34	1550_TSS10_part3_F	regA2	<u>CCGGTACGGGAGAT</u> CTCCCTATAGTGAGTCGTATTAATCAAAGGCGTATGTCCACTGGG
oCW_35	1551_TSS10_part3_R	regA2	ACTGATAGGGAGATCTAGGTGTAAGGAAAACTGAATGAGA
oCW_36	1552_TSS10_part4_F	regA3	<u>CCGGTACGGGAGAT</u> CTCCCTATAGTGAGTCGTATTAATCAGGCCCTTTGGTTACCCACTT
oCW_37	1553_TSS10_part4_R	regA3	ACTGATAGGGAGATCTTCCCACCGTGAGTTAAACCAG
oCW_38	1554_TSS10_part5_F	regA4	<u>CCGGTACGGGAGAT</u> CTCCCTATAGTGAGTCGTATTAATCACAAAATTATGGTGCACGTAC GGT
oCW_39	1555_TSS10_part5_R	regA4	ACTGATAGGGAGATCTTACTTCTAGGTGGGGCTCCC
oCW_40	1556_TSS10_part6_F	regA5	<u>CCGGTACGGGAGAT</u> CTCCCTATAGTGAGTCGTATTAATCAGCAACGTCTCCTTCGCTCTT
oCW_41	1557_TSS10_part6_R	regA5	ACTGATAGGGAGATCTAGCCAAGAGGTTTGTGGTTCA
oCW_42	1558_TSS10_part7_F	regA6	<u>CCGGTACGGGAGAT</u> CTCCCTATAGTGAGTCGTATTAATCAGGCACCCAACTGCTGAAAG
oCW_43	1559_TSS10_part7_R	regA6	ACTGATAGGGAGATCTGCAAAATGCATACGCTCGGT
oCW_44	1546_TSS10_part1_F	regA7	<u>CCGGTACGGGAGAT</u> CTCCCTATAGTGAGTCGTATTAATCACGAGGCTTTTGCTAAGAGGG T
oCW_45	1547_TSS10_part1_R	regA7	ACTGATAGGGAGATCTAGCTCAGCAGTAATAAGGGTCA
oCW_46	1612_TSS10_1/2_inv_F	regA2rc	<u>CCGGTACGGGAGAT</u> CTCCCTATAGTGAGTCGTATTAATCAAGATCTAGGTGTAAGGAAAA CTGAATGAGA
oCW_47	1613_TSS10_1/2_inv_R	regA2rc	ACTGATAGGGAGATCTAAAGGCGTATGTCCACTGGG
oCW_48	1405_pCW26_TSS16_1_F	regB	<u>CCGGTACGGGAGAT</u> CTCCCTATAGTGAGTCGTATTAATCACCCGAAAGTGATGAGGGGA G
oCW_49	1545_TSS16_part5_R	regB	ACTGATAGGGAGATCTAAAACAATATTTTTCTTCGTCAGCGT
oCW_50	1536_TSS16_part1_F	regB1	<u>CCGGTACGGGAGAT</u> CTCCCTATAGTGAGTCGTATTAATCACCCGAAAGTGATGAGGGGA G
oCW_51	1537_TSS16_part1_R	regB1	ACTGATAGGGAGATCTCTTATCTGTCCACCAATAGAGTATTTTC
oCW_52	1538_TSS16_part2_F	regB2	<u>CCGGTACGGGAGAT</u> CTCCCTATAGTGAGTCGTATTAATCAGAACGCTAACCCCTCCTCG

2. Materials and methods

Oligo name	Lab internal name	Purpose	DNA sequence 5'-3'
oCW_53	1539_TSS16_part2_R	regB2	<u>ACTGATAGGGAGATCTGTGGGACAAACACGGTCACT</u>
oCW_54	1540_TSS16_part3_F	regB3	<u>CCGGTACGGGAGATCTCCCTATAGTGAGTCGTATTAATCAGGAGTACTAAAGTGCTGCGGA</u>
oCW_55	1541_TSS16_part3_R	regB3	<u>ACTGATAGGGAGATCTGCAAAGAAGACCATTTCGTCAACA</u>
oCW_56	1542_TSS16_part4_F	regB4	<u>CCGGTACGGGAGATCTCCCTATAGTGAGTCGTATTAATCATACTCCTTTTGCTTGCGGCG</u>
oCW_57	1543_TSS16_part4_R	regB4	<u>ACTGATAGGGAGATCTCACTCATTTTCATAACCGGTCCG</u>
oCW_58	1544_TSS16_part5_F	regB5	<u>CCGGTACGGGAGATCTCCCTATAGTGAGTCGTATTAATCATTTCGTTCTTTGATCAAAGTGACGT</u>
oCW_59	1545_TSS16_part5_R	regB5	<u>ACTGATAGGGAGATCTAAAACAATATTTTTCTTCGTCAGCGT</u>
oCW_60	1652_TSS16_1/1revcomp_F	regB1rc	<u>GTATTAATCAAGATCTCTTATCTGTCCACCAATAGAGTATTTTC</u>
oCW_61	1653_TSS16_1/1revcomp_R	regB1rc	<u>ACTGATAGGGAGATCTCCCGGAAAGTGATGAGGGAG</u>
oCW_62	1313_pCW27_frgmt2_F	pCW27v2 UHR	<u>ATTAATGCAGCTCGAGAAGTCAGAAGGGGAAAGCGG</u>
oCW_63	1560_pCW27v2_pre_R	pCW27v2 UHR	<u>TCCGGATAGGCTTAAGGTTGTACTGGGAGAGGGTGC</u>
oCW_64	1561_pCW27v2_F	pCW27v2 DHR	<u>AGGCATGCAAGCTAGCCGCGCGCATCTCAAATCTAC</u>
oCW_65	1562_pCW27v2_R	pCW27v2 DHR	<u>AGAGGATCTGGCTAGGCCGGCCGCGGGTGCTTGCCTTTCATCAC</u>
oCW_66	1314_pCW28_frgmt2_F	pCW28v2 UHR	<u>ATTAATGCAGCTCGAGCTTTCAGCAAGCACGCAGAG</u>
oCW_67	1563_pCW28v2_pre_R	pCW28v2 UHR	<u>TCCGGATAGGCTTAAGCGGGAAGAGGTGGTGAACTT</u>
oCW_68	1564_pCW28v2_F	pCW28v2 DHR	<u>GGCATGCAAGCTAGCGGTTCCCTGTGCATAATTCGC</u>
oCW_69	1565_pCW28v2_R	pCW28v2 DHR	<u>AGAGGATCTGGCTAGGCCGGCCGCTCCTCCCATAAATGTACAGCTCG</u>
oCW_70	1771_BLE_F	Phleomycin	<u>ACAGAACAATTTTTGGCCACACAACCCGGTGTTAGGATCTCCGAGGCCTTTAGTCCTGCTCTCGGCC</u>
oCW_71	1775_BLE_R2	Phleomycin	<u>AAGCTCTAGAAGTATGGCCAAGTTGACCAGTGC</u>
oCW_72	1646_TSSmotifs1_R	pCW24v2_GT_210_nt verification	ACACACAACACCACCGACAA
oCW_73	1647_TSSmotifs2_R	pCW24v2_GT_416_nt verification	AGGGGAGGGGGACACAC
oCW_74	1648_TSSmotifs1_revcomp_F	pCW24v2_GT_210_nt_rc	<u>GTATTAATCAAGATCTAACAACAACCCTCCACACAC</u>
oCW_75	1649_TSSmotifs1_revcomp_R	pCW24v2_GT_210_nt_rc	<u>CCGGCCGTAGGCGCGCCGTGTGTGGTGCTTTTT</u>
oCW_76	1650_TSSmotifs1/2_revcomp_F	pCW24v2_GT_416_nt_rc	<u>GTATTAATCAAGATCTACAAACAACACAAAAA</u>

2. Materials and methods

Oligo name	Lab internal name	Purpose	DNA sequence 5'-3'
oCW_77	1651_TSSmotifs1/2_revcomp_R	pCW24v2_GT_416_nt_rc	<u>TGATAGGGAGATCCGGCCGGCCGTGTGTGTGGTGCTTTT</u>
oCW_78	1662_pAK1_rRNAR	pCW24v4 verification	GGGAGATCTACCGTACGCCGTAAGCGCTACTTTTACTGC
oCW_79	1797_pCW24v3_TSSmotif1/2_R	pCW24v4_GT_416_nt verification	AACTCAGTACTCAGGCCGGCCACAAACAAACACAA
oCW_80	1792_pCW24v3_TSSmotif1rc_F	pCW24v4_GT_210_nt_rc	<u>GTATTAATCAAGATCTTCCCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGATAACAACAACCCTCCACACAC</u>
oCW_81	1793_pCW24v3_TSSmotif1rc_R	pCW24v4_GT_210_nt_rc	<u>CCGGCCGTAGGCGCGCCGTGTGTGTGGTGCTTTTT</u>
oCW_82	1794_pCW24v3_TSSmotif1/2rc_F	pCW24v4_GT_416_nt_rc	<u>GTATTAATCAAGATCTTCCCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGATTCTACAAACAAACACAAAA</u>
oCW_83	1795_pCW24v3_TSSmotif1/2rc_R	pCW24v4_GT_416_nt_rc	<u>CCGGCCGTAGGCGCGCCCGTGTGTGTGGTGCTTTT</u>
oCW_84	1763_invHyg_F	Hygromycin inversion	<u>TTGTTAGCAGCATTAAATCCCGTACCGGGGGCACA</u>
oCW_85	1764_invHyg_R	Hygromycin inversion	<u>ACTGATAGGGAGATCTTTGCAGAATACTGCATAGATAACAAACGC</u>
oCW_86	1765_pCW37_UR_F	pCW37 UHR	<u>ATTAATGCAGCTCGAGCCCTCTGTTTTCACCTCCTCC</u>
oCW_87	1766_pCW37_UR_R	pCW37 UHR	<u>TCCGGATAGGCTTAAGAACGAGGAGGAGGGCAAAG</u>
oCW_88	1767_pCW37_DR_F	pCW37 DHR	<u>AGGCATGCAAGCTAGCAATTTCTCCACCTGTTTCACACT</u>
oCW_89	1768_pCW37_DR_R	pCW37 DHR	<u>ICTGGCTAGGCGGCCGCTCACTTGCTTTCACTTCTTCACTTC</u>
oCW_90	1778_optPolyY_plus	anneals to long polyY tract	<u>ACATGTTCTCGTCCCGGGTTTTTTTTCTTTTTTTTTCTTTTTTTTTTATAGACTTCAATTACACCAAAAAGTAAAATTCACAAGCTTGAATTCCTT</u>
oCW_91	1779_optPolyY_minus	anneals to long polyY tract	<u>AAGGAATTCGAAGCTTGTGAATTTTACTTTTTGGTGTAATTGAAGTCTATAAAAAAAAAAAGAAAAAAAAAGAAAAAAAAACCCGGGACGAGAACATGT</u>
oCW_92	1782_noPolyY_plus	anneals to no polyY tract	<u>ACATGTTCTCGTCCCGGGATAGACTTCAATTACACCAAAAAGTAAAATTCACAAGCTTGAATTCCTT</u>
oCW_93	1783_noPolyY_minus	anneals to no polyY tract	<u>AAGGAATTCGAAGCTTGTGAATTTTACTTTTTGGTGTAATTGAAGTCTATCCCGGGACGAGAACATGT</u>

Table 2.4 List of gBlocks used in this study.

Underlined sequences indicate InFusion overhangs.

Sequence name	Purpose	DNA sequence 5'-3'
pCW24v2 background		
GT_210_nt	insertion of short GT-rich sequence	<u>GTATTAATCAAGATCT</u> GTGTGTGTGGTGCTTTTTTCGCTTTTTTGTGTGTGGGGCGAAGAAAATGTTTGTGTTCTTTCTC CCGTGTGTGTGCTTCCCCCTTTGTGCGTGCGTAGGGGGAGAGTTCCCCCTTTGGGGGGAAACTGTGTGTGGGGTTTGTGTTG TGTGGGTGCGGGGGGGGAAACTTTGTTTTGTCGGTGGTGTGTGTGTGGAGGGTTGTTGTTGGCGCGCCTACGGCCGGCC <u>GATCTCCCTATCAGT</u>
GT_206_nt	extension of the short GT-rich sequence	<u>GTTGTTGTTGGCGCG</u> CCGCGTGTGTGTAATGTTTTTGGGGGGGAGTTATTTTTGGTGGGTGGTGGTGTGTGTGTGCG TGCGTGTTTTGTTTTGGGGGGGGTTTTCCCCCTTTGTTGCTGCTGTTTTTTGTTCCCCCTCCCCTGTGTGTGTGTGTGTTTGT GCCCTTTTCTTGTGTTGTGTGCCCCCTCCCCTGTTTTTGTGTTTGTGTTGTTGGCCGG <u>CCGATCTCCCTATCA</u>
pCW24v4 background		
rRNA_promoter	insertion of the rRNA promoter	<u>AGTCGTATTAATCAGGTACCT</u> CCCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGATTAGCTTTCCACCCAGCGCGGG TGCATTCTGGCTCTTATATATACTTATTGTCATGACAGAGTATATTGACTGTGTTGATAAGGGACGGGTAAGTATTGAAG AGCCGATGCTTTTGACATGTTAGATATAATATGTTTTATTGTAAAGTCAATACAACACACAATAGGATAATAATGATAAAGTTA AAAAAGTATATATAGTAATAGAAATATATCTTATATAGGAAAGATTAAGCAGTAAAAGTAGCGCTTACGGCGTACGGTCCCTG AGTACTGAGTTTAAACATGTTCTCGTCCC <u>GGGCTGCACGCGCCT</u>
no_promoter	no-promoter control	<u>AGTCGTATTAATCAGGTACCT</u> CCCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGATTAGCCCTGAGTACTGAGTTTAA CATGTTCTCGTCCC <u>GGGCTGCACGCGCCT</u>
GT_210_nt	insertion of short GT-rich sequence	<u>GCCCCCGGTACGGG</u> AGATCCCCTATAGTGAAGTCGTATTAATCAAGATCTTCCCTATCAGTGATAGAGATCTCCCTATCAGTG ATAGAGATGTGTGTGTGGTGCTTTTTTCGCTTTTTTGTGTGTGGGGCGAAGAAAATGTTTGTGTTCTTTCTCCCGTGTG TGTGCTTCCCCCTTTGTGCGTGCGTAGGGGGAGAGTTCCCCCTTTGGGGGGAAACTGTGTGTGGGGTTTGTGTTGTGGGT GCGGGGGGGGAAACTTTGTTTTGTCGGTGGTGTGTTGTGTGTGGAGGGTTGTTGTTGGCGCGCCTACGGCCGGCCTGAGTAC TGAGTTTAAACATGTTCTCGTCCC <u>GGGCTGCACGCGCCT</u>
GT_206_nt	extension of the short GT-rich sequence	<u>GTTGTTGTTGGCGCG</u> CCGCGTGTGTGTAATGTTTTTGGGGGGGAGTTATTTTTGGTGGGTGGTGGTGTGTGTGTGCG TGCGTGTTTTGTTTTGGGGGGGGTTTTCCCCCTTTGTTGCTGCTGTTTTTTGTTCCCCCTCCCCTGTGTGTGTGTGTGTTTGT GCCCTTTTCTTGTGTTGTGTGCCCCCTCCCCTGTTTTTGTGTTTGTGTTTGTGGCCGG <u>CCTGAGTACTGAGTT</u>

2.3 Trypanosome cell culture and analysis

2.3.1 Trypanosome growth

HMI-11 medium, pH 7.5

0.3% NaHCO₃ (Roth), 0.0136% Hypoxanthine (Sigma Aldrich), 0.018% L-Cysteine (Sigma Aldrich), 0.003% Bathocuprione disulfonic acid·Na₂ salt (Serva), 1% PenStrep (Thermo Scientific), 0.0005% β-Mercaptoethanol, 1.76% IMDM (Thermo Scientific), 10% FCS (Sigma Aldrich)

T. brucei bloodstream form cell lines (Table 2.5 and Table 2.6) were cultured in HMI-11 medium at 37 °C and 5% CO₂ to a density of 0.8-1.8 x 10⁶ cells/ml. HMI-11 medium contains the same ingredients as HMI-9 (Hirumi and Hirumi, 1994), but lacks serum plus. Additionally, we dispensed with adding thymidine. Where appropriate, the following drug concentrations were used: 2 μg/ml G418 (Roth), 5 μg/ml hygromycin (Roth), 0.1 μg/ml puromycin (Sigma Aldrich), 25 μg/ml blasticidin (Invivogen), 1 μg/ml doxycycline (Sigma Aldrich), 50 μg/ml ganciclovir (Invivogen).

Table 2.5 List of parental *T. brucei* cell lines used in this study.

Abbreviations: BF, bloodstream form; MITat1.2, Molteno Institute Trypanozoon Antigen Type 1.2; SM, single marker; TETR, tetracycline repressor, T7RNAP, T7 RNA polymerase; Δ, deletion; NEO, aminoglycoside phosphotransferase; BSD, blasticidin S deaminase; PAC, puromycin N-acetyltransferase; HYG, hygromycin phosphotransferase; G418, neomycin; Blas, blasticidin; Puro, puromycin; Hygro, hygromycin.

Name	Genotype	Constructs	Selection	Reference
Wt	BF Lister 427, MITat1.2, clone 221a	-	-	(Cross, 1975)
SM	BF Lister 427, MITat1.2, clone 221a, <i>TETR T7RNAP NEO</i>	pHD328, pLew114hyg5'	G418	(Wirtz <i>et al.</i> , 1999)
BFJEL43	SM, RRNA::Ty1- <i>H2B.V BSD</i> , Δ <i>H2B.V::PAC</i> Δ <i>H2B.V::HYG</i>	pJEL92, pJEL74, pJEL75	Blas, Puro, Hygro	(Lowell <i>et al.</i> , 2005)
BFJEL25	SM, Δ <i>H3V::PAC</i> Δ <i>H3V::HYG</i>	pJEL89, pJEL38	G418, Puro, Hygro	(Lowell and Cross, 2004)

Table 2.6 List of transgenic *T. brucei* cell lines generated in this study.

Abbreviations: SM, single marker; ψ 221, VSG pseudogene of active 221 BES; RLUC, Renilla luciferase; RRNA, ribosomal RNA locus; BSD, blasticidin S deaminase; BLE, phleomycin resistance gene; PAC, puromycin N-acetyltransferase; HYG, hygromycin phosphotransferase; G418, neomycin; Blas, blasticidin; Phleo, phleomycin; Puro, puromycin; Hygro, hygromycin.

Name	Genotype	Constructs	Selection	Reference
SM Rluc	SM, ψ 221::RLUC BSD	pCJ25ARluc	G418, Blas	C Wedel
SM Ty1-RPB9	SM Rluc, 2xTy1-RPB9 BLE, 2xTy1-RPB9 PAC	pPOTv3_TY-RPB9_Phleo, pPOTv3_TY-RPB9_Puro	G418, Blas, Phleo, Puro	C Wedel
SM Ty1-H2A.Z H2A.Z ⁻	SM, RRNA::Ty1-H2A.Z BLE, Δ H2A.Z::HYG	pLEW111_TY1-H2A.Z, pyrFEKO-HYG_H2A.Z	G418, Phleo, Hygro	C Wedel
SM Ty1-H2A.Z H2A.Z ^{-/-}	SM Ty1-H2A.Z H2A.Z ⁻ , Δ H2A.Z::PAC	pyrFEKO-PUR_H2A.Z, pLEW100Cre-EP1	G418, Phleo	C Wedel
pCW24v2	SM Rluc, Tb427_01_v4:283,211::RRNAPROM::FLUC HYG	pCW24v2	G418, Blas, Hygro	C Wedel, R Derr
pCW24v2-p	SM Rluc, Tb427_01_v4:283,211::FLUC HYG	pCW24v2-p	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc regA	SM Rluc, Tb427_01_v4:283,211::regA::FLUC HYG	pCW24v2_regA	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc regB	SM Rluc, Tb427_01_v4:283,211::regB::FLUC HYG	pCW24v2_regB	G418, Blas, Hygro	C Wedel, R Derr
pCW27v2	SM Rluc, Tb427_09_v4:1,067,648::RRNAPROM::FLUC HYG	pCW27v2	G418, Blas, Hygro	C Wedel
pCW27v2-p	SM Rluc, Tb427_09_v4:1,067,648::FLUC HYG	pCW27v2-p	G418, Blas, Hygro	C Wedel
pCW27v2 regA	SM Rluc, Tb427_09_v4:1,067,648::regA::FLUC HYG	pCW27v2_regA	G418, Blas, Hygro	C Wedel
pCW27v2 regB	SM Rluc, Tb427_09_v4:1,067,648::regB::FLUC HYG	pCW27v2_regA	G418, Blas, Hygro	C Wedel
pCW28v2	SM Rluc, Tb427_10_v5:1,927,049::RRNAPROM::FLUC HYG	pCW28v2	G418, Blas, Hygro	C Wedel
pCW28v2-p	SM Rluc, Tb427_10_v5:1,927,049::FLUC HYG	pCW28v2-p	G418, Blas, Hygro	C Wedel
pCW28v2 regA	SM Rluc, Tb427_10_v5:1,927,049::regA::FLUC HYG	pCW28v2_regA	G418, Blas, Hygro	C Wedel
pCW28v2 regB	SM Rluc, Tb427_10_v5:1,927,049::regB::FLUC HYG	pCW28v2_regA	G418, Blas, Hygro	C Wedel
SM Rluc regA1	SM, Tb427_01_v4:283,211::regA1::FLUC HYG	pCW24v2_regA1	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc regA2	SM, Tb427_01_v4:283,211::regA2::FLUC HYG	pCW24v2_regA2	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc regA3	SM, Tb427_01_v4:283,211::regA3::FLUC HYG	pCW24v2_regA3	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc regA4	SM, Tb427_01_v4:283,211::regA4::FLUC HYG	pCW24v2_regA4	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc regA5	SM, Tb427_01_v4:283,211::regA5::FLUC HYG	pCW24v2_regA5	G418, Blas, Hygro	C Wedel, R Derr

2. Materials and methods

Name	Genotype	Constructs	Selection	Reference
SM Rluc regA6	SM, Tb427_01_v4:283,211::regA6::FLUC HYG	pCW24v2_regA6	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc regA7	SM, Tb427_01_v4:283,211::regA7::FLUC HYG	pCW24v2_regA7	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc regA2rc	SM, Tb427_01_v4:283,211::regA2rc::FLUC HYG	pCW24v2_regA2rc	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc regB1	SM Rluc, Tb427_01_v4:283,211::regB1::FLUC HYG	pCW24v2_regB1	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc regB2	SM Rluc, Tb427_01_v4:283,211::regB2::FLUC HYG	pCW24v2_regB2	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc regB3	SM Rluc, Tb427_01_v4:283,211::regB3::FLUC HYG	pCW24v2_regB3	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc regB4	SM Rluc, Tb427_01_v4:283,211::regB4::FLUC HYG	pCW24v2_regB4	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc regB5	SM Rluc, Tb427_01_v4:283,211::regB5::FLUC HYG	pCW24v2_regB5	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc regB1rc	SM Rluc, Tb427_01_v4:283,211::regB1rc::FLUC HYG	pCW24v2_regB1rc	G418, Blas, Hygro	C Wedel, R Derr
SM Rluc GT_210_nt	SM Rluc, Tb427_01_v4:283,211::GT_210_nt::FLUC HYG	pCW24v2_GT_210_nt	G418, Blas, Hygro	C Wedel
SM Rluc GT_210_nt_rc	SM Rluc, Tb427_01_v4:283,211::GT_210_nt_rc::FLUC HYG	pCW24v2_GT_210_nt_rc	G418, Blas, Hygro	C Wedel
SM Rluc GT_416_nt	SM Rluc, Tb427_01_v4:283,211::GT_416_nt::FLUC HYG	pCW24v2_GT_416_nt	G418, Blas, Hygro	C Wedel
SM Rluc GT_416_nt_rc	SM Rluc, Tb427_01_v4:283,211::GT_416_nt_rc::FLUC HYG	pCW24v2_GT_416_nt_rc	G418, Blas, Hygro	C Wedel
pCW27v3 GT_416_nt	BFJEL25, Tb427_09_v4:1,067,648::GT_416_nt::FLUC BLE	pCW27v3_GT_416_nt	G418, Blas, Hygro	C Wedel
pCW28v3 GT_416_nt	BFJEL25, Tb427_10_v5:1,927,049::GT_416_nt::FLUC BLE	pCW28v3_GT_416_nt	G418, Blas, Hygro	C Wedel
pCW37	SM Rluc, Tb427_01_v4:500,640::FLUC HYG	pCW37	G418, Blas, Hygro	C Wedel
pCW29 clone 1	SM Rluc, RRNA::FLUC HYG	pLEW100v5_HYG	G418, Blas, Hygro	C Wedel
pCW29 clone 2	SM Rluc, RRNA::FLUC HYG	pLEW100v5_HYG	G418, Blas, Hygro	C Wedel
pCW29 clone 3	SM Rluc, RRNA::FLUC HYG	pLEW100v5_HYG	G418, Blas, Hygro	C Wedel
pCW29 clone 4	SM Rluc, RRNA::FLUC HYG	pLEW100v5_HYG	G418, Blas, Hygro	C Wedel
pCW24v4	SM Rluc, Tb427_01_v4:283,211::RRNAPROM::FLUC HYG	pCW24v4	G418, Blas, Hygro	C Wedel
pCW24v4-p	SM Rluc, Tb427_01_v4:283,211::FLUC HYG	pCW24v4-p	G418, Blas, Hygro	C Wedel
pCW24v4 GT_210_nt	SM Rluc, Tb427_01_v4:283,211::GT_210_nt::FLUC HYG	pCW24v4_GT_210_nt	G418, Blas, Hygro	C Wedel
pCW24v4 GT_210_nt_rc	SM Rluc, Tb427_01_v4:283,211::GT_210_nt_rc::FLUC HYG	pCW24v4_GT_210_nt_rc	G418, Blas, Hygro	C Wedel
pCW24v4 GT_416_nt	SM Rluc, Tb427_01_v4:283,211::GT_416_nt::FLUC HYG	pCW24v4_GT_416_nt	G418, Blas, Hygro	C Wedel
pCW24v4 GT_416_nt_rc	SM Rluc, Tb427_01_v4:283,211::GT_416_nt_rc::FLUC HYG	pCW24v4_GT_416_nt_rc	G418, Blas, Hygro	C Wedel
pCW24v2 long polyY	SM Rluc, Tb427_01_v4:283,211::GT_210_nt::FLUC HYG	pCW24v2_longPolyY	G418, Blas, Hygro	C Wedel
pCW24v2 no polyY	SM Rluc, Tb427_01_v4:283,211::GT_210_nt::FLUC HYG	pCW24v2_noPolyY	G418, Blas, Hygro	C Wedel

2.3.2 Cryo stock preparation and reconstitution

To preserve *T. brucei* cell lines, cryo stocks were generated. 2×10^6 BF cells were harvested by centrifugation at 800 xg for 10 min and the supernatant was removed. The cell pellet was resuspended in 1 ml of HMI-11 medium containing 10% glycerol and transferred into a cryo stock vial. The cryo stocks were frozen using a Mr. Frosty freezing container (Thermo Scientific) to ensure slow freezing (-1 °C/min) and stored at -80 °C. For long term storage, the vials were transferred to liquid nitrogen.

To reconstitute preserved *T. brucei* cell lines, the cryo vial was thawed in a 37 °C water bath for 1 min and the cells were immediately transferred into 9 ml of pre-warmed HMI-11 medium. Where appropriate, drugs were added to the in chapter 2.3.1 mentioned concentrations after 6 h.

2.3.3 Stable transfection of *T. brucei* cells

Transfection buffer (Schumann Burkard *et al.*, 2011)

90 mM Na₂HPO₄, 5 mM KCl, 0.15 mM CaCl₂, 50 mM HEPES-NaOH pH 7.3

Stable transfections in BF have been performed as described by Scahill *et al.*, 2008. In brief, 3×10^7 BF cells were harvested by centrifugation at 800 xg for 10 min and the supernatant was removed. The cell pellet was resuspended in 200 μ l of cold transfection buffer and transferred to a pre-chilled electroporation cuvette (2 mm, 400 μ l, VWR). Upon addition of 10 μ g of linearized plasmid the cells were transfected using the Nucleofector 2b (Lonza) and the program X-001. The cells were immediately transferred to 30 ml of pre-warmed HMI-11 medium and diluted in a 1:10 and 1:100 ratio. The three dilutions (undiluted, 1:10, 1:100) were spread on 24 well-plates (1 ml/well) and incubated at 37 °C and 5% CO₂. After 6 h the respective drugs were added to select for cells, which had integrated the construct successfully into the genome.

2.3.4 Transient transfection of *T. brucei* cells

In this study, the cell line SM Ty1-H2A.Z H2A.Z^{-/-} was transiently transfected with the circular plasmid pLEW100Cre-EP1 to remove the floxed resistance genes and the *Herpes simplex* virus thymidine kinase *HSVTK* introduced to knock out both endogenous *H2A.Z* alleles. The transfection was performed as described in chapter 2.3.3 with minor changes. In brief,

3×10^7 BF cells were harvested and transfected in 250 μ l transfection buffer with 20 μ g of circular pLEW100Cre-EP1 and transferred in HMI-11 medium supplemented with 0.3 μ g/ml doxycycline to induce the expression of CRE recombinase. The cells were diluted 1:10, 1:100, 1:1,000 and 1:10,000 in HMI-11 medium supplemented with 0.3 μ g/ml doxycycline and each dilution was spread on 24 well-plates. After 6 h ganciclovir was added (50 μ g/ml final concentration) to select for the loss of *HSVTK*. In parallel, a mock transfection was performed without the addition of pLEW100Cre-EP1 to ensure sufficient killing of *HSVTK*-positive cells.

2.3.5 Isolation of genomic DNA

The Phusion Human Specimen Direct PCR Kit (Thermo Scientific) was used to isolate genomic DNA from *T. brucei* BF cells. 1×10^6 cells were harvested by centrifugation at 1,500 x g for 10 min at RT and the supernatant was removed. The cell pellet was resuspended in 20 μ l of dilution buffer and 0.5 μ l DNA release additive. The reaction was mixed, incubated for 5 min at RT and boiled for 2 min at 98 °C. The cell debris was separated by centrifugation at 2,000 xg for 5 min at RT and 1 μ l of the supernatant was used as template for PCR (see chapter 2.1.1).

2.3.6 Isolation of RNA, cDNA synthesis and qPCR analysis

Total RNA of 5×10^7 *T. brucei* BF cells was extracted using the NucleoSpin RNA kit from Macherey&Nagel according to the instructions of the manufacturer. One μ g of total RNA was used to synthesize cDNA using M-MLV Reverse Transcriptase (Thermo Scientific) according to the instructions of the manufacturer. To analyse the expression of specific genes qPCR was performed with 1 μ l of cDNA (diluted 1:8 in dH₂O), the iTaq Universal SYBR Green Supermix (Bio-Rad) and the following cycling conditions: 95 °C/30 sec, 35 cycles (95 °C/5 sec – 60 °C/30 sec), melting curve 65 °C-95 °C (0.5 °C increment 2-5 sec/step) in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Primers were chosen to have an optimal melting temperature at 60 °C and to amplify 90-150 bp. The result was analysed using the double delta Ct method based on the expression of TERT (telomerase reverse transcriptase, Tb11.01.1950), since it is described to be the most suitable reference gene in the *T. brucei* genome (Brenndorfer and Boshart, 2010).

2.3.7 Dual-Luciferase assay

Luciferase activities were measured with the Dual-Glo Luciferase Assay System from Promega. *T. brucei* BF cells were grown to a maximum density of 1.8×10^6 cells/ml and

5×10^6 cells were harvested, washed with 1x PBS and resuspended in 200 μ L 1x PBS. In a 96-well plate 50 μ l cell suspension (1.25×10^6 cells) and 50 μ l Dual-Glo Luciferase Reagent were mixed and after 10 min firefly luciferase activity was measured for 1 sec in a Victorlight Luminometer. 50 μ l Dual-Glo Stop&Glo Reagent were added and after 10 min renilla luciferase activity was measured. All measurements were performed at least in duplicates. To normalize for differences in cell number, firefly luciferase activity was normalized with renilla luciferase activity.

2.3.8 Fluorescence microscopy

PBS (phosphate buffered saline), pH 7.4

2.7 nM KCl, 2 mM KH_2PO_4 , 10 mM Na_2HPO_4 , 137 mM NaCl

PBG

PBS, 0.2% cold water fish skin gelatin (Sigma Aldrich), 0.5% BSA

A total of 2×10^7 *T. brucei* BF cells were harvested by centrifugation at 800 xg for 10 min at RT and fixed in HMI-11 medium supplemented with 2% formaldehyde for 5 min. The cells were washed three times with PBS and allowed to settle down on (2-aminopropyl)triethoxysilane-coated cover slips for 30 min. Unsettled cells were removed by washing with PBS. The cells were permeabilized for 5 min in 100 μ l of PBS containing 0.2% NP-40 and washed three times with PBS. Upon two blocking steps with 100 μ l PBG for 10 min the cells were incubated with the primary antibody diluted in PBG for 1.5 h. Subsequently, the cells were washed four times with PBG for 5 min and incubated with the secondary antibody coupled to a fluorophore for 1.5 h in the dark. After an additional washing step, the cells were mounted in Vectashield DAPI (Vecta Laboratories) on a microscope slide and the edges of the cover slip were sealed with nail polish. The samples were incubated over night at RT in the dark and stored at 4 °C. The samples were analyzed using a Leica DMI6000B microscope. Images were captured using Leica Application Suite and false colored using Fiji.

2.4 Biochemical methods

2.4.1 Western blot

Lysis buffer

1 ml 4x Laemmli sample buffer, 3 ml RIPA buffer, 8 μ l 1 M DTT

4x Laemmli sample buffer

40% glycerol, 0.02% bromphenolblue, 8% SDS, 250 mM Tris pH 6.5

RIPA buffer

50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40, 0.25% Na-Deoxycholate, 0.1% SDS

PBS (phosphate buffered saline), pH 7.4

2.7 mM KCl, 2 mM KH₂PO₄, 10 mM Na₂HPO₄, 137 mM NaCl

Per condition 2×10^6 *T. brucei* cells were harvested and lysed using lysis buffer. The cell lysates were separated on a 15% SDS-PAGE gel and transferred onto a nitrocellulose membrane using a semi-dry blotting method. The membranes were stained with 0.5% AmidoBlack staining solution in 10% acetic acid to verify the transfer, destained in destaining solution (25% isopropanol, 10% acetic acid) and subsequently blocked for 1 h in PBS-T (PBS + 0.1% Tween20) containing 3% BSA. The membranes were hybridized with the primary antibody diluted in PBS-T containing 1% milk over night at 4 °C. Primary antibodies were detected using secondary antibodies (α -mouse or α -rabbit, GE Healthcare) conjugated to horseradish peroxidase diluted in PBS-T containing 1% milk for 1.5 h at RT and Pierce ECL Western Blotting Substrate (Thermo Scientific) using a Luminescent Image Analyzer (GE Healthcare).

2.4.2 Antibody production

To perform MNase-ChIP-seq experiments against the endogenous histone variants H2A.Z (Tb427.07.6360) and H2B.V (Tb427tmp.02.5250), polyclonal antibodies were generated by immunizing three rabbits per protein with the respective antigen (α H2A.Z #1-3, α H2B.V #1.3). MNase-ChIP-seq is performed under native conditions. Under non-denaturing conditions, histones are only partially accessible to proteins since the majority of the protein is incorporated into nucleosomes. Only the highly post-translationally modified N-terminal tail

(30-40 aa) protrudes from the nucleus. These modifications are trypanosome specific and the reason why commercially available antibodies cannot be used. Thus, a 22-aa and 18-aa sequence from H2A.Z (DDAVPQAPLVGGVAMSPEQAS) and H2B.V (SSSSRKKSGARRGKKQQ) were selected based on unpublished PTM analysis data from Johannes Thürich to synthesize peptides used for immunization. Only completely unmodified aa were considered to avoid selection of proteins containing either modified or unmodified aa only (Appendix Figure 7.1A). The peptide was synthesized and immunization was performed by Pineda Antikörper Service. The synthesized peptide was covalently coupled to KLH (Keyhole limpet haemocyanin) to increase immunogenicity and HPLC purified. Prior to immunization preimmune sera of 12 rabbits were tested for cross reactions via western blot (data not shown). The immunization was performed by Pineda Antikörper Service according to the following 145-day protocol (Appendix Figure 7.1B): day 1, intradermal preimmunization with peptide in CFA (Complete Freund's adjuvant); day 20/30/40/61/75/90/115, subcutaneous boosts with peptide in IFA (Incomplete Freund's adjuvant). On day 61/90/120 test bleeds were analyzed for antibody production via western blot (Appendix Figure 7.1C). Based on the result of the first test bleed on day 61 the whole serum of α H2A.Z #1 was harvested. α H2A.Z #2 died on day 81. The whole serum of α H2A.Z #3 and α H2B.V #1-3 was harvested on day 145. The retrieved antisera were stored in 5-ml aliquots at -80 °C and α H2A.Z #1 75 d was affinity purified to evaluate different elution strategies (high salt, low pH, high pH, Appendix Figure 7.1D). The antisera α H2A.Z #1 75 d and α H2B.V #2 145 d were affinity purified by applying low pH conditions during elution. The purified antibodies were verified via western blot (Appendix Figure 7.1E), immunofluorescence microscopy (Appendix Figure 7.1F) and MNase-ChIP-seq (data not shown), where signals from α H2A.Z could be obtained using all methods and from α H2B.V using western blot only.

2.4.3 Antibody affinity purification

PBS (phosphate buffered saline), pH 7.4

2.7 nM KCl, 2 mM KH₂PO₄, 10 mM Na₂HPO₄, 137 mM NaCl

The polyclonal antisera α H2A.Z and α H2B.V produced in chapter 2.4.1 were affinity purified using SulphoLink Coupling Gel (Thermo Scientific) with the respective immobilized peptide as described elsewhere (Harlow *et al.*). In brief, 1 mg peptide was covalently coupled to the beaded SulphoLink Coupling Gel in a column. To block nonspecific binding sites of the gel the coupled beads were incubated with 50 mM L-cysteine. Specific antibodies from antisera were

purified by repetitive application of the serum onto the peptide-coupled gel column. For α H2A.Z three strategies were tested to elute the specific antibodies in 6 fractions each from the column: high salt using 3 M potassium chloride, low pH using 0.1 M glycine pH 2.5 and high pH using 0.1 M glycine pH 10. The low pH elution turned out to be the most suitable (Appendix Figure 7.1D). The elution solution was dialyzed against PBS over night at 4 °C and concentrated the next day using Spectra/Gel® Absorbent (Spectrumlabs.com) for several hours at RT. The concentrations were determined using NanoDrop (Thermo Fisher).

2.5 Next-generation sequencing methods

2.5.1 MNase-ChIP-seq

Formaldehyde solution

50 mM HEPES-KOH pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 11%
Formaldehyde

Permeabilization buffer

100 mM KCl, 10 mM Tris pH 8.0, 25 mM EDTA

Protease inhibitors

1.46 μ M Pepstatin A, 4.7 μ M Leupeptin, 1 mM PMSF, 1 mM TLCK

NP-S buffer

0.5 mM Spermidine, 0.075% IGEPAL, 50 mM NaCl, 10 mM Tris-HCl pH 7.5, 5 mM
MgCl₂, 1 mM CaCl₂

RIPA buffer

50 mM HEPES-KOH pH 7.5, 500 mM LiCl, 1 mM EDTA, 1.0% NP-40, 0.7% Na-
Deoxycholate

Elution buffer

50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1.0% SDS

Immunoprecipitation of mono-nucleosomes was performed as described previously (Wedel and Siegel, 2017). A total of 2×10^8 cells at a density of $0.8-1.5 \times 10^6$ cells/ml were harvested, resuspended in 30 ml of HMI-11 medium and cross-linked using 4 ml of formaldehyde solution.

Depending on the abundance and distribution of the target protein along the genome the amount of starting material can be decreased to 0.8×10^8 cells to have enough DNA to prepare the sequencing libraries. When immunoprecipitating epitope-tagged proteins, it is important to consider that solvent-accessible lysine residues provide the most reactive functional groups in native proteins for formaldehyde (Hoffman *et al.*, 2015). Thus, the presence of lysine residues within epitope tags may cause the tag to be cross-linked to the core protein, leaving the tag inaccessible to antibodies. The number of lysine residues within commonly used tags varies from FLAG containing 2 lysine residues over Myc containing 1 to Ty1 not containing any lysine. The cross-link reaction is quenched by adding an excess amount of glycine (108 mM final conc.) as it serves as additional reaction partner for formaldehyde. Cell lysis was performed using 200 μ M digitonin in permeabilization buffer containing protease inhibitors. The cells were washed in NP-S buffer containing protease inhibitors and EDTA and chromatin was fragmented using MNase. A critical step is to remove all traces of EDTA by several washing steps prior to MNase treatment, since MNase activity is dependent on Ca^{2+} , which is chelated by EDTA. Because the activity of MNase varies among vendors, batch and incubation time, it is necessary to titrate the amount of applied MNase to prevent over- or under-digestion (Figure 2.1A). The goal is to digest the chromatin to ~95 % mono-nucleosomes (147 bp) and ~5 % di-nucleosomes (~350 bp). Additionally, MNase activity varies between uncross-linked and cross-linked chromatin (Figure 2.1B). The MNase activity was stopped by adding EDTA and soluble nucleosomes were separated by centrifugation. The pellet was resuspended in NP-S buffer and to increase the yield of the experiment, chromatin remaining in the pellet was solubilized by mild sonification. After centrifugation both supernatants were pooled and an aliquot was separated, which was used as input control. This material was treated like the eluate after immunoprecipitation and served as size control and to normalize for technical artefacts introduced during the library construction, sequencing or computational analysis. Immunoprecipitation of nucleosomal DNA was performed using Dynabeads® M-280 Sheep anti-Rabbit IgG (Invitrogen) coupled to polyclonal H3 rabbit-antiserum (Gassen *et al.*, 2012) at 4 °C for 2 h or coupled to my custom-made polyclonal affinity-purified H2A.Z rabbit-antibody (this study) at RT for 30 min in the presence of 0.05% SDS to remove uncross-linked DNA fragments. It is possible to use monoclonal as well as polyclonal antibodies, but it is important that the antibody is specific and able to bind its epitope under cross-linked conditions. This can be tested by western blotting and performing immunofluorescence microscopy prior to MNase-ChIP-seq, respectively. Bound material was washed several times with RIPA buffer and eluted in elution buffer. Cross-links of the input and the ChIPed sample were reversed over night at 65 °C in presence of 300 mM NaCl and mono-nucleosomal DNA was purified using the

NucleoSpin Gel and PCR Clean-up Kit and Buffer NTB (Macherey&Nagel). The DNA was quantified using the Qubit dsDNA HS Assay Kit and the Qubit 2.0 Fluorometer (Thermo Scientific).

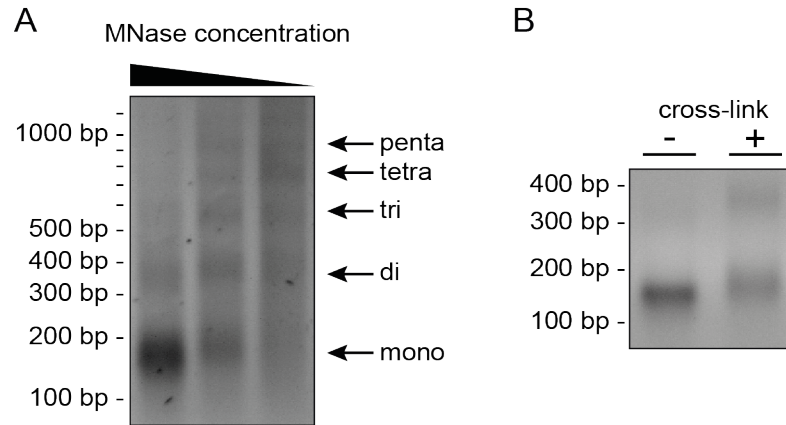


Figure 2.1 MNase activity varies with concentration and substrate.

(A) Titration of MNase concentration. The chromatin of 2×10^6 cells was treated with 1 U, 0.25 U and 0.0625 U, respectively for 5 min at 25 °C. 300 ng of each nucleosomal DNA was run on a 2 % agarose gel (left lane: 1 U, central lane: 0.25 U, right lane: 0.0625 U). (B) 2 % agarose gel with uncross-linked (-) and cross-linked (+) chromatin treated with 2 U MNase for 10 min at 25°C. A different MNase batch was used between the experiments depicted in (A) and (B).

2.5.2 Library construction

ChIP-seq libraries were constructed as described in (Nguyen *et al.*, 2014) and (Wedel and Siegel, 2017). A total of 35 ng of immunoprecipitated DNA or 35 ng of input DNA were end-repaired, A-tailed, barcoded and Y-shaped adapters were ligated to each end of the DNA molecules (Figure 2.2A). Each Y-shaped adapter is composed of a universal sequence and a unique sequence containing the barcode sequence (Figure 2.2B). Both sequences share a partial homology region basepairing during annealing prior to the ligation to the DNA fragments. The included barcode allows pooling of several samples and reduces sequencing cost. The chimeric DNA fragments were purified and size selected to exclude unbound adapters using Agencourt AMPure XP beads (Beckman Coulter). During 5 cycles of PCR with KAPA Hot Start ReadyMix (KAPA Biosystems) Y-shaped adapters were converted into double-stranded DNA. After an additional round of purification with Agencourt AMPure XP beads adapter dimers were removed by size selection from an 2.5% agarose gel to increase the yield of usable sequencing reads. The size selected DNA was extracted from the gel using the NucleoSpin Gel and PCR Clean-up Kit (Macherey&Nagel). In a last step the library was amplified by PCR. It is important not to over-amplify the library as this will lead to the formation

of duplexes that may cause problems during sequencing. In addition, the more cycles the PCR runs, the more likely it is that amplification biases are introduced. Thus, as few cycles as possible should be used. The number of cycles is dependent on the amount of starting material. When starting with 10-35 ng 9 and 11 cycles have been used. After the amplification the library is analyzed on an agarose gel and if duplexes are visible the PCR should be repeated with fewer cycles. The DNA was quantified in duplicates using the Qubit dsDNA HS Assay Kit and the Qubit 2.0 Fluorometer (Thermo Scientific).

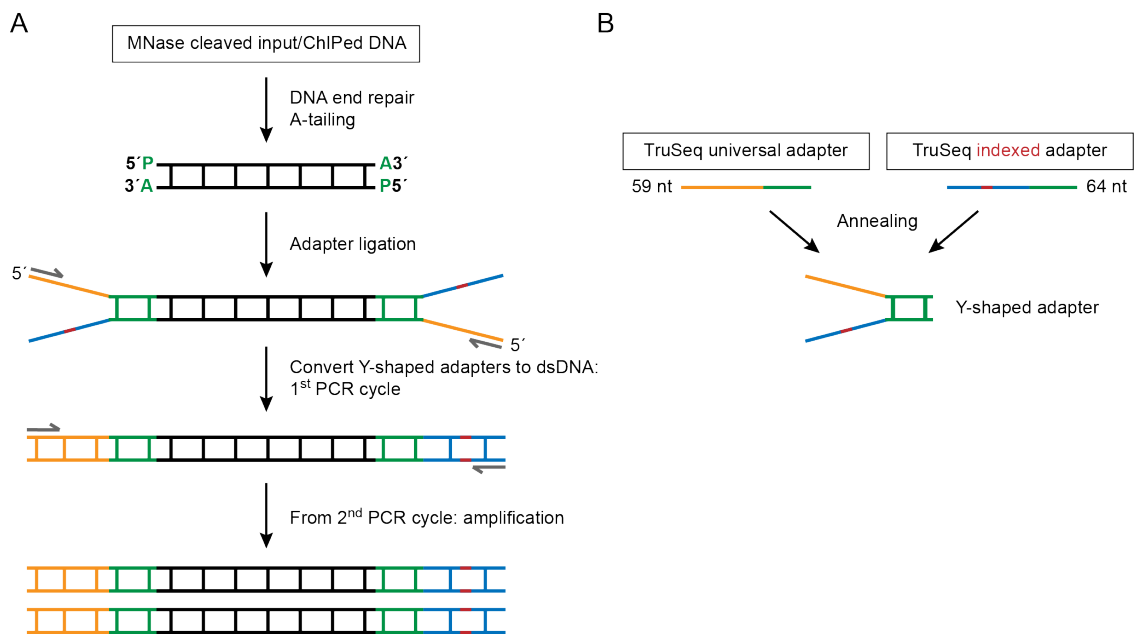


Figure 2.2 NGS library construction.

(A) Schematic overview of the construction of a sequencing library using Y-shaped adapters. Fragmented DNA is blunt-ended and A-tailed. Y-shaped adapters are ligated to each end of each DNA molecule and are converted to dsDNA by PCR. The library is amplified during an additional PCR. (B) Y-shaped adapters are generated by annealing of a TruSeq universal adapter and a TruSeq indexed adapter with partial sequence homologies (green).

2.5.3 Library quantification and sequencing

An exact quantification of the sequencing library is essential to obtain a high yield of sequencing reads of high quality. First, the total DNA concentration of the library was determined using a Qubit and to determine the concentration of DNA containing adapters a qPCR is performed. When Qubit and qPCR results differed, qPCR values have been used. For each library the molarity was calculated based on the size of the containing DNA molecules and adjusted for all samples, e.g. to 2 nM and pooled where applicable. Using the KAPA SYBR Fast Universal qPCR Kit (KAPA Biosystems) a 1:500 and 1:1000 dilution of the pool or the sample was quantified in duplicates. Based on the calculated molarity a 0.5, 1 or 2 nM

sequencing pool was generated. The ratio of the samples within the pool determines the number of reads for each sample. The number of sequencing reads needed for a successful analysis depends on the experimental scope and the size of the genome. Proteins strongly enriched at specific sites can easily be identified at relatively low coverage while subtle patterns of widely distributed proteins may only be revealed at high sequencing depth, illustrated in Figure 2.3. The *T. brucei* genome is about 30 Mb in size. At this size, I found that 1×10^6 reads were sufficient to detect strong peaks (Figure 2.3A), while the precise mapping of nucleosomes required a much larger number of reads (Figure 2.3B). Unlike ChIPed DNA, the input material consists of fragmented DNA from the entire genome. Thus, to obtain sufficient coverage to allow for normalization at the sites of interest, input material should be sequenced at least to the same depth as the ChIPed DNA. The DNA was chemically denatured following the instructions of Illumina and diluted to a final concentration of 1.8 pM and mixed with 1 % PhiX control to monitor the overall sequencing performance. Datasets generated within this study were sequenced using an Illumina HiSeq 2500 (generating 100 bp-reads) or NextSeq 500 platform (generating 76 bp-reads). All datasets generated and used in this study and information about sequencing and subsequent analysis are listed in Appendix Table 7.3 and Appendix Table 7.4.

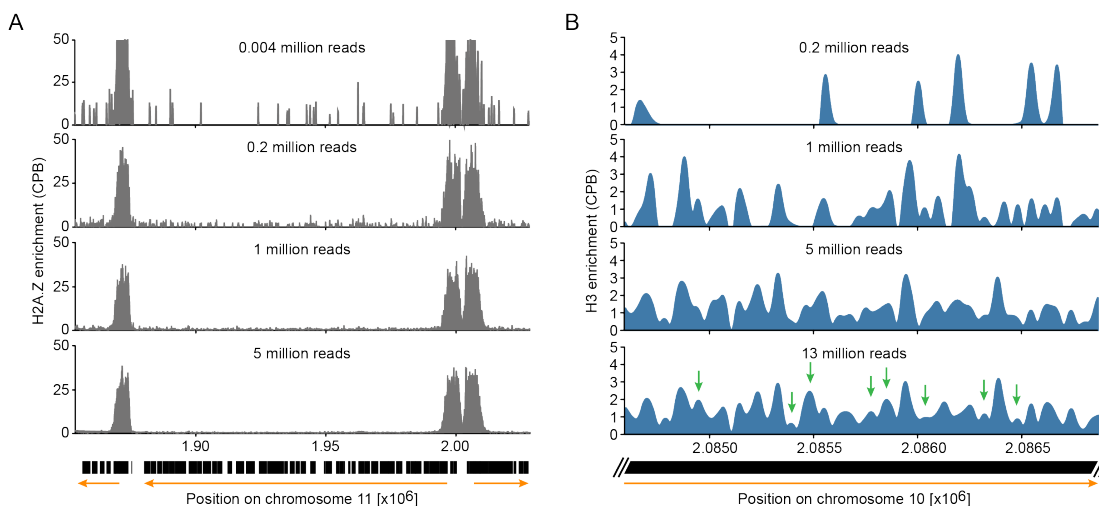


Figure 2.3 The number of required reads depends on the scope of the experiment.

(A) The distribution of the histone variant H2A.Z was mapped using MNase-ChIP-seq as representative for detection of strong peaks (Wedel and Siegel, 2017). The different panels illustrate the effect of sequencing depth. Reads are mapped to the Tb427v24 genome using bowtie2 and processed using samtools and COVERnant. Black boxes and orange arrows indicate ORFs and direction of transcription, respectively. (B) The distribution of H3 was mapped using MNase-ChIP-seq to generate a nucleosome occupancy map (Wedel and Siegel, 2017). The different panels illustrate the effect of sequencing depth. Green arrows highlight nucleosome positions becoming clearer with higher sequencing depth. Reads are mapped to the Tb427v8 genome using bowtie1 and processed according to the nucwave pipeline (Quintales *et al.*, 2015).

2.5.4 Computational analysis

For each sequencing run two .fastq files for each, the forward and the reverse strand are obtained. To remove the adapter sequences from the sequencing reads the adapter-trimming algorithm cutadapt (Martin, 2011) with the following command was used:

```
$ cutadapt -a <ADAPTER_FWD> -A <ADAPTER_REV> -o <out.1.fastq> -p  
<out.2.fastq> <reads.1.fastq> <reads.2.fastq>
```

To map the trimmed sequencing reads I used the short-read mapping software bowtie2 (Langmead and Salzberg, 2012) and downloaded the sequence (.fasta file) and the annotation file (.gff file) of the respective reference genome from EuPathDB (Aurrecochea *et al.*, 2013). In general, any short-read mapping software can be used. An alternative to bowtie2 is BMA-mem (Li and Durbin, 2010), which is excellent in terms of accuracy and it allows analysis of low complexity regions. I preferred using bowtie2, because of its speed while using a personal computer. Bowtie2 requires the installation of three additional softwares: samtools (Li *et al.*, 2009a), GNU curses library (<http://www.gnu.org/software/ncurses>), ZLib compression library (<http://zlib.net>). As a first step, I indexed the genome prior to sequencing read mapping using the following command:

```
$ bowtie2-build <reference genome.fasta> <output prefix>
```

Next, I generated a .sam file to map the sequencing reads.

```
$ bowtie2 -t --local -x <indexed genome prefix> -1 <read 1> -2  
<read 2> > [output.sam]
```

Depending on the scope of the experiment, I removed reads mapping to several locations in the genome, i.e. non-unique reads. Those can be removed by using samtools and the following command:

```
$ samtools view -Sh <input.sam> | grep -v "XS:i:" > <output.sam>
```

To visualize the sequencing data in a browser I converted the .sam file to an indexed .sorted.bam file. Therefore, I first converted the .sam file to a .bam file. After sorting and

indexing the .bam file I visualized the sequencing data using the IGV browser (Robinson *et al.*, 2011).

```
$ samtools view -bh <input.sam> > <output.bam>
```

```
$ samtools sort <input.bam> > <output.sorted.bam>
```

```
$ samtools index <output.sorted.bam>
```

For the analysis of my sequencing data I exclusively used our home-made pipeline COVERnant, which was developed by Konrad Förstner and the Siegel lab (Wedel *et al.*, 2017). To generate coverage files that are normalized to the sequencing depths (i.e. counts per billion reads, CPB) or to determine the ratio of ChIPed DNA/input signal, I used COVERnants subcommand 'ratio', which generates .wig files. Here, a window size (ws) and a step size (ss) can be set to define the area size for which the number of reads is counted and to define the step size by which the window is moved along the genome.

```
$ covernant ratio --paired_end --keep_zero_coverage <ws> <ss> -o  
<output> --denominator <denominator.sorted.bam> --numerator  
<numerator.sorted.bam>
```

To average the sequencing data across multiple regions I used COVERnants subcommand 'extract'. I generated a .csv file containing the coordinates of the regions for which the meta plot should be generated.

```
$ covernant extract <ws> <ss> --output_prefix <output prefix>  
--flip_reverse_strand <input.wig> <coordinates.csv>
```

This generates two files. The _matrix.csv file contains the raw values for the regions chosen in the input .csv file. The median, mean and standard deviation calculated from these data are saved in the _matrix.csv file and can be visualized using a statistics software, i.e. GraphPad Prism. All datasets generated and used in this study and information about sequencing and subsequent analysis are listed in Appendix Table 7.3 and Appendix Table 7.4.

2.5.5 Ty1-RPB9-ChIP-seq

The RNA pol II-ChIP was performed in cell lines, in which both endogenous alleles of RPB9, an essential subunit of the RNA pol II complex, contained a Ty1 epitope tag. Except for minor changes, the ChIP was performed as described in chapter 2.5.1. In brief, 3×10^8 cells were harvested, formaldehyde cross-linked and permeabilized using digitonin. After centrifugation, the pellet was resuspended in 600 μ l of NP-S buffer and sonicated for 50 cycles at low strength in a 15 ml tube. After centrifugation, the supernatant was transferred to a new microcentrifuge tube and 60 μ l were separated as input. Immunoprecipitation of DNA bound by RBP9 was performed using Dynabeads Protein G (Invitrogen) coupled to a BB2 antibody (Bastin *et al.*, 1996) at 4 °C over night and under presence of 300 mM NaCl. Bound material was washed, cross-links were reversed and immunoprecipitated DNA was purified. Sequencing libraries were constructed as described in chapter 2.5.2 and sequenced on an Illumina NextSeq 500.

2.5.6 5'PPP-RNA-seq

Small RNA (< 200 nt) were purified from 5×10^7 cells using a combination of miRNeasy Mini (Qiagen) and RNeasy MinElute Cleanup (Qiagen) following the instructions of the manufacturer. A total of 2 μ g of small RNA were treated with 1 unit of Terminator 5'-Phosphate-Dependent Exonuclease (TEX, Epicentre) in 1x Terminator Reaction Buffer B for 30 min at 42 °C to remove 5'-monophosphate RNA. Next, the RNA was purified and one half (+5' Polyphosphatase) was treated with 20 units RNA 5' Polyphosphatase to convert 5'-triphosphate RNA to 5'-monophosphate RNA. The second half (-5' Polyphosphatase) was left untreated and served as a control. The sequencing library was constructed using the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England BioLabs). Briefly, a 3'-adapter was ligated to the RNA and the RT primer was hybridized to the 3'-adapter generating a double-stranded 3'-end. This increases the efficiency of the 5'-adapter ligation as it preferentially binds single-stranded molecules. In addition, the RT primer also hybridizes to unbound 3'-adapters minimizing the formation of adaptor-dimers during 5'-adapter ligation. In a next step, RNA was converted into cDNA by extension of the first strand from the RT primer during a PCR. In a second PCR, molecules with an adaptor at each end were enriched, barcodes were incorporated and the amount of the library was increased. A size selection step ensured that only small RNA molecules were included in the final library. From the control sample a library was generated, as well. Here, small transcripts containing a 5'-triphosphate

were excluded from the library, since the 5'-adapter requires a 5'-monophosphate. The samples were sequenced on an Illumina NextSeq 500.

2.6 Data generated in this study and source code availability

All sequencing data generated in this study have been deposited in NCBI's Gene Expression Omnibus (Edgar *et al.*, 2002) and are accessible through GEO Series accession number GSE98061 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98061>). All datasets generated and used in this study and information about sequencing and subsequent analysis are listed in Appendix Table 7.3 and Appendix Table 7.4.

2.7 Software

The following table lists software used to analyze the data generated in this study.

Table 2.7 Software used in this study.

Software	Reference
cutadapt v1.14	(Martin, 2011); http://cutadapt.readthedocs.io/en/stable/index.html
IGV browser v2.4.10	(Robinson <i>et al.</i> , 2011); http://software.broadinstitute.org/software/igv
GNU curses library v6.0	http://www.gnu.org/software/ncurses
ZLib compression library v1.2.10	http://zlib.net
bowtie2 v2.1.0	(Langmead and Salzberg, 2012); https://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.1.0/
samtools v0.1.19e44428cd	(Li <i>et al.</i> , 2009a); http://samtools.sourceforge.net
COVERnant v0.3.0	(Wedel <i>et al.</i> , 2017); https://github.com/konrad/COVERnant
GraphPad Prism v5.0b	http://www.graphpad.com
Leica Application Suite Advanced Fluorescence (LAS AF) v3.3	https://www.leicabiosystems.com/de/arbeitsablauf/klinische-mikroskopie/dokumentation-und-software/
Fiji v1.0	(Schindelin <i>et al.</i> , 2012); http://imagej.net/Fiji

3

Characterization of RNA pol II transcription start regions

3.1	The RNA pol II subunit RPB9 is enriched at the 5'-end of TSRs	79
3.2	Transcription initiates ~200 bp upstream of RPB9 enrichment.....	80
3.3	Chromatin structure around transcription start regions.....	82
3.4	Concluding remarks	86

So far, transcription start regions (TSRs) in *T. brucei* have been demonstrated to be marked by the co-localization of the histone variants H2A.Z and H2B.V, the histone modifications H3K4me3 and H4K10ac, the bromodomain protein BDF3 and primary transcripts (Siegel *et al.*, 2009; Kolev *et al.*, 2010; Wright *et al.*, 2010). To further characterize TSRs, I aimed to identify sites of RNA pol II-specific transcription initiation following a two-pronged approach. First, I determined the genome-wide distribution of RNA pol II enrichment by ChIP-seq. Subsequently, I mapped the distribution of small primary transcripts carrying a 5'-triphosphate that can be used to identify sites of transcription initiation. Overlapping enrichment of both, RNA pol II and small primary transcripts, would identify those sites where RNA pol II-mediated transcription initiates. Furthermore, to investigate whether the chromatin structure within TSRs is distinct to that of the remaining genome, I conducted MNase-ChIP-seq to investigate the chromatin accessibility within and around TSRs by mapping TSR-nucleosomes.

3.1 The RNA pol II subunit RPB9 is enriched at the 5'-end of TSRs

To determine the genome-wide RNA pol II enrichment by ChIP-seq, I generated a cell line expressing both endogenous alleles of the essential RNA pol II complex subunit RPB9 fused N-terminally to two Ty1 epitope tags using the pPOTv3 system (Dean *et al.*, 2015). The cell line was confirmed by integration PCR and western blot (Figure 3.1A and B). The ChIP-seq experiment was performed with chromatin fragmented by sonication and Ty1-RPB9-bound DNA was immunoprecipitated using the BB2 antibody (Bastin *et al.*, 1996). Mapping of the RPB9-ChIP-seq data revealed a strong enrichment of the RNA pol II complex at the 5'-end of each individual TSR, which is marked by H2A.Z-enrichment. The distribution of RPB9 and H2A.Z enrichment at a representative region on chromosome 9 is depicted in the left panel of Figure 3.1C. When averaging RPB9 enrichment for all non-divergent TSRs in the *T. brucei* genome, the RPB9 enrichment can be localized to a defined region of ~2 kb in size at the 5'-end of H2A.Z enrichment (Figure 3.1C, right panel). This enrichment has been demonstrated in other eukaryotes, as well, and is due to promoter-proximal pausing upon transcription initiation (Adelman and Lis, 2012). Thus, I hypothesize that the identified sites of RNA pol II complex enrichment described here mark sites of RNA pol II pausing, just downstream of RNA pol II transcription initiation sites.

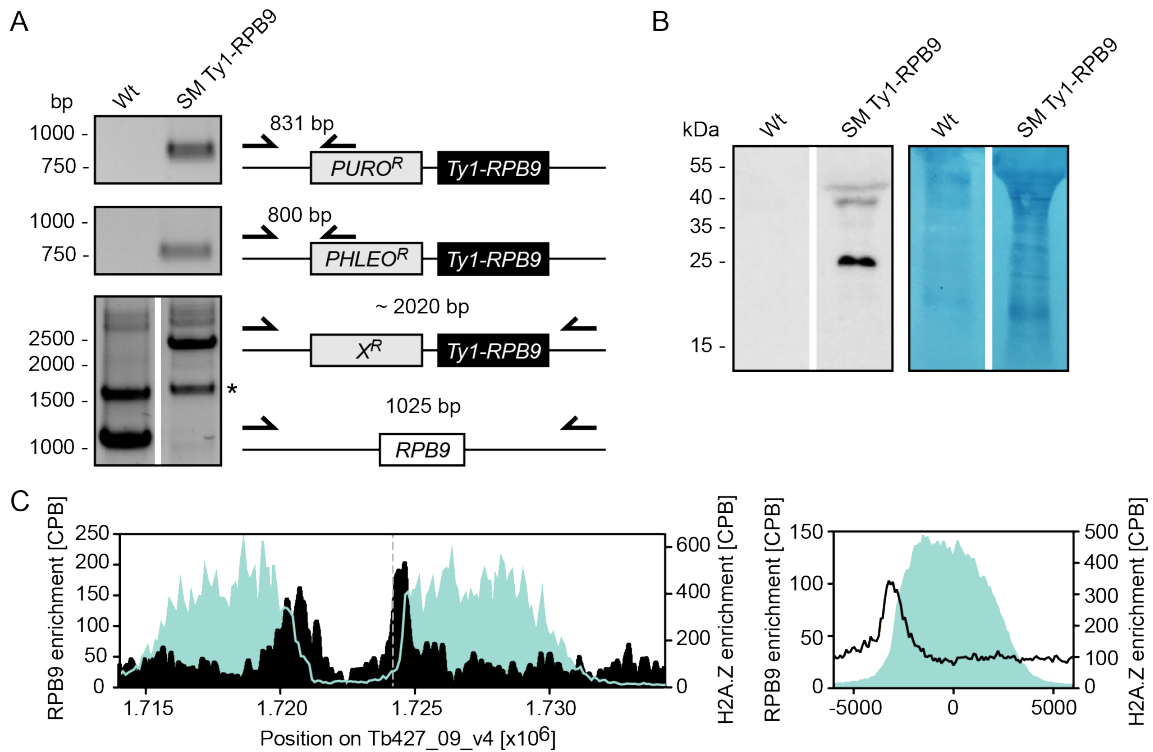


Figure 3.1 RPB9 is enriched at the 5'-end of H2A.Z enrichment.

(A) Verification of the Ty1-RPB9 cell line by integration PCR. Primers bind upstream of the integration site and in the resistance gene upon transfection with pPOTv3_TY-RPB9_Puro (top panel) and pPOTv3_TY-RPB9_Phleo (middle panel) and up- and downstream of the Wt or tagged allele (bottom panel; X^R, respective resistance gene). Unspecific PCR products are marked by an asterisk. (B) Verification of the Ty1-RPB9 cell line by western blot using the BB2 antibody to detect Ty1-RPB9 (left panel). The amido black stained nitrocellulose is shown as loading control (right panel). (C) RPB9 ChIP-seq data (black) shown for the same representative region as in Figure 3.2B (left panels) and averaged for all 57 non-divergent TSRs (right panel). A dashed line added where RNA pol II enrichment is considered to start. H2A.Z enrichment determined by MNase-ChIP-seq is shown in cyan.

3.2 Transcription initiates ~200 bp upstream of RPB9 enrichment

To test whether sites of RNA pol II complex enrichment indeed mark pausing sites downstream of transcription initiation, I mapped the distribution of small primary transcripts carrying a 5'-triphosphate. Only primary transcripts contain a 5'-triphosphate since the triphosphate is processed during RNA maturation. The sequencing of the primary transcriptome of *T. brucei* has been successfully performed by Kolev and colleagues (Kolev *et al.*, 2010). In an effort to improve the localization of transcription initiation sites, I used a different approach. The experimental outline is illustrated in Figure 3.2A. Kolev and colleagues isolated total RNA that was enriched for polyadenylated RNA and depleted the sample of rRNA by Terminator 5'-Phosphate-Dependent Exonuclease (TEX) treatment, which digests 5'-monophosphate-containing RNA. In comparison, I exclusively isolated RNA that was smaller than 200 nt from total RNA. Thus, I enriched the sample for short newly transcribed RNA, avoided to isolate long and abundant rRNA and to introduce a bias for already processed RNA. Subsequently, I

treated the sample with TEX to remove remnant 5′-monophosphate-containing RNA. To account for undigested 5′-monophosphate-containing RNA contaminants, the sample was split, and one half was treated with 5′-polyphosphatase to reduce the 5′-triphosphate to a 5′-monophosphate. The other half was left untreated and was used to normalize for undigested 5′-monophosphate-containing RNA contaminants in the sample used to identify transcription initiation sites upon sequencing.

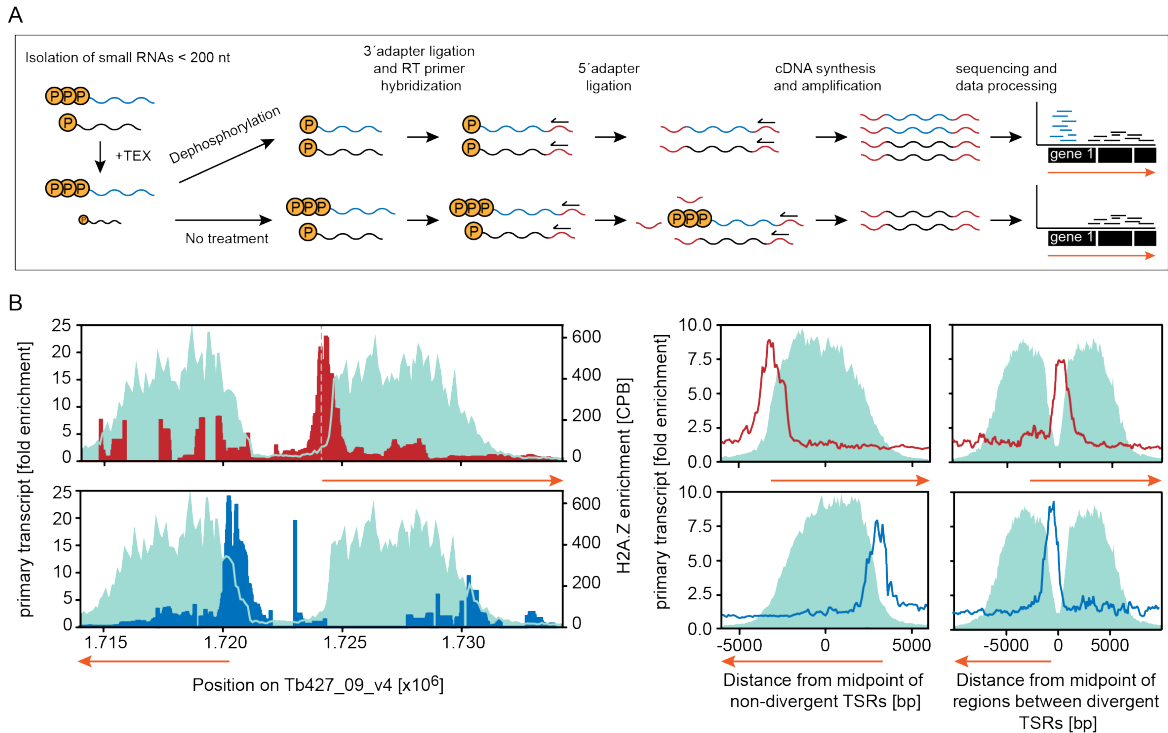


Figure 3.2 Identification of transcription initiation sites by mapping small primary transcripts.

(A) Schematic outline of small 5′-triphosphate-RNA-seq. Small total RNA < 200 nt was purified from *T. brucei*. Treatment of the RNA with Terminator 5′-Phosphate-Dependent Exonuclease (TEX) degraded 5′-monophosphate-containing RNA (black) and thus enriched for primary transcripts containing a 5′-triphosphate (blue). To normalize for undigested monophosphate-containing RNA contaminants, the sample was split and libraries were prepared from 5′-Polyphosphatase-treated and untreated material. CPB, counts per billion reads. (B) Strand-specific mapping of small primary transcripts shown for a representative divergent TSR on chromosome 9 (left panel) and averaged for 27/30 non-divergent and 71 divergent TSRs (middle and right panels). The grey dashed line indicates the start of RPB9 enrichment (Figure 3.1B). Transcripts derived from the bottom strand and the top strand are shown in red and blue, respectively. H2A.Z enrichment determined by MNase-ChIP-seq is shown in cyan. Orange arrows indicate the direction of transcription.

The sequencing reads were mapped strand-specifically to the *T. brucei* genome and normalized as described above using COVERnant (see chapter 2.5.4). The enrichment of H2A.Z (determined in chapter 3.3.2) was used as reference for TSR location. The mapping revealed a strong enrichment of primary transcripts at the 5′-end of H2A.Z enrichment. The distribution of primary transcripts and H2A.Z enrichment at a representative region on chromosome 9 is depicted in the left panel of Figure 3.2B. When averaging primary transcript

enrichment for all TSRs in the *T. brucei* genome, the enrichment can be localized to a defined region of ~2 kb in size at the 5'-end of H2A.Z enrichment (Figure 3.2B, right panel). The enrichment of 5'-triphosphate-containing RNA resembled the pattern of RPB9. When comparing the peaks of both, RNA pol II complex and primary transcripts, the former seemed to be enriched just downstream of the latter (compare dashed lines in Figure 3.1C and Figure 3.2B). This finding suggested that RNA pol II transcription pauses 100-200 bp downstream of its initiation, similar to what has been observed in metazoans (Adelman and Lis, 2012). In addition, the data indicate a strong directionality of transcription initiation as the ratio of sense to antisense primary transcripts is 4:1 for both, the top and the bottom strand. Taken together, these findings suggest that the sites enriched in RNA pol II indeed mark sites of directional RNA pol II-mediated transcription initiation and that those sites locate to the 5'-end of TSRs. A list of RNA pol II transcription initiation sites can be found in Appendix Table 7.1.

3.3 Chromatin structure around transcription start regions

In *T. brucei*, TSRs are marked by H2A.Z-containing nucleosomes. Previous studies performed in different organisms, including *T. brucei*, have suggested that nucleosomes containing H2A.Z are less stable compared to canonical nucleosomes (Suto *et al.*, 2000; Abbott *et al.*, 2001; Zhang *et al.*, 2005; Jin and Felsenfeld, 2007; Siegel *et al.*, 2009). Thus, H2A.Z-containing nucleosomes contribute to a more open chromatin structure and an increased accessibility of DNA to proteins. Accessibility of DNA packaged into chromatin can be investigated by digestion of the chromatin with micrococcal nuclease (MNase), which specifically digests DNA that is not protected by histones. Thus, to investigate whether the DNA within TSRs in *T. brucei* is more accessible to proteins, I established an MNase-ChIP-seq approach for *T. brucei* (Wedel and Siegel, 2017).

3.3.1 MNase-ChIP-seq – A high-resolution method to investigate chromatin accessibility by mapping nucleosome positioning

The key steps of the established MNase-ChIP-seq protocol are illustrated in Figure 3.3A. In brief, DNA-protein interactions are formaldehyde cross-linked and the cells are permeabilized. MNase is used to fragment the chromatin, such that the majority of the chromatin is digested to mono-nucleosomes and that a small population of di-nucleosomal DNA remains to avoid over-digestion (Figure 3.3B). To enrich for nucleosomal DNA, nucleosomes are immunoprecipitated using a histone-specific antibody. Following immunoprecipitation, the

cross-links are reversed, the DNA is purified and used to generate Illumina sequencer-compatible libraries, which are subjected to next-generation sequencing. MNase preferentially digests DNA, which is not protected by proteins, such as the linker DNA between nucleosomes. The resulting digestion products, the nucleosomal DNA, is on average 147 bp in size. The size can differ depending on how tight the DNA is wrapped around the histone octamer (illustrated in Figure 3.3C). Chromatin containing loosely bound nucleosomes is more accessible to MNase and yields on average shorter MNase cleavage products (< 147 bp) than more compact chromatin (Weiner *et al.*, 2010). As a result, when nucleosomal DNA is sequenced and mapped back to the genome nucleosome occupancy maps can be generated revealing nucleosome positioning (Cole *et al.*, 2012). When exclusively sub-nucleosomal DNA < 147 bp is analyzed the position of loosely bound nucleosomes can be identified revealing regions of increased MNase sensitivity, reflecting the accessibility to proteins. A detailed protocol can be found in chapters 2.5.1-2.5.4.

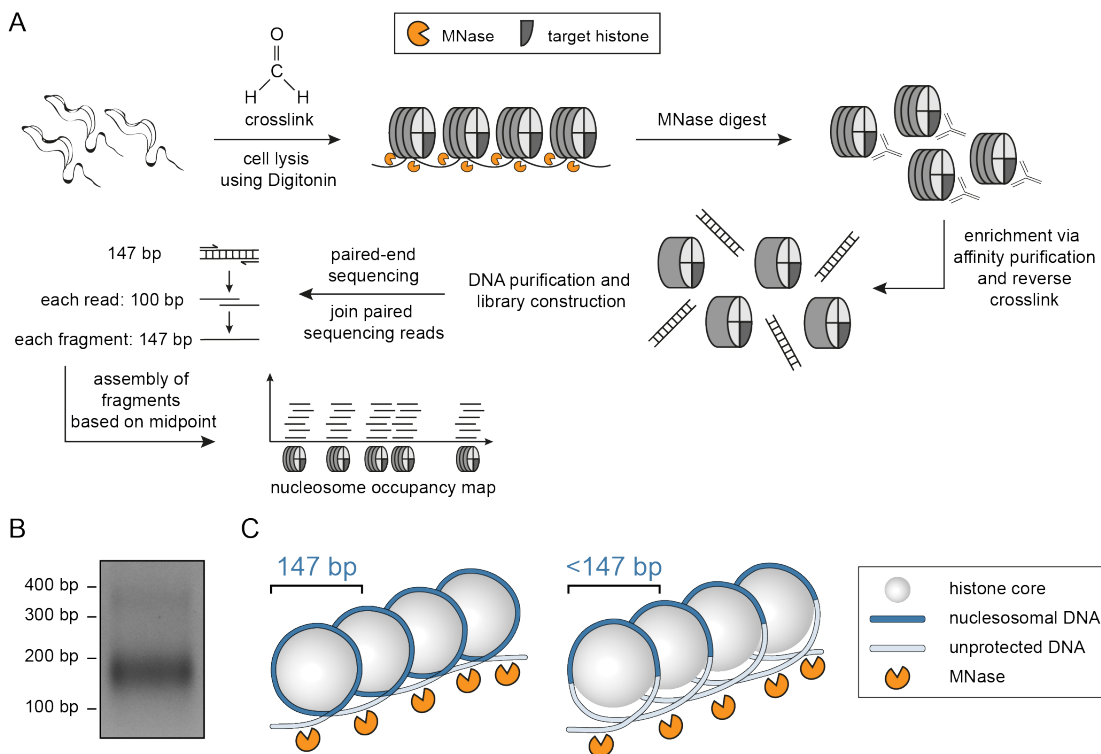


Figure 3.3 Establishment of a high-resolution MNase-ChIP-seq protocol for *T. brucei*.

(A) Outline of MNase-ChIP-seq. *T. brucei* cells are formaldehyde cross-linked, permeabilized and chromatin is digested into mono-nucleosomes using MNase. Nucleosomes are immunoprecipitated using a histone-specific antibody. After the reversal of cross-links, the nucleosomal DNA is purified and paired-end-sequenced. The sequencing reads are joined to fragments and assembled according to their midpoints generating nucleosome occupancy maps. (B) 2% agarose gel with 100 ng of mono-nucleosomal DNA after MNase digest. (C) Illustration of nucleosomal DNA length after MNase digestion of average bound nucleosomes (left panel) and loosely bound nucleosomes (right panel).

To establish the MNase-ChIP-seq protocol outlined above I joined steps from several previously published protocols. Cross-linking was performed as described by Lee et al. (Lee *et al.*, 2006). To generate, isolate and immunoprecipitate mono-nucleosomes I combined a protocol for the permeabilization of *T. brucei* cells (Lowell *et al.*, 2005) with a strategy to obtain and pull-down a high yield of mono-nucleosomes that had been used in *S. cerevisiae* (Wal and Pugh, 2012). For the washing and purification of the ChIPed DNA I included the protocol developed by Siegel and co-workers (Siegel *et al.*, 2009). To construct ChIP-seq libraries I used a protocol published by Ethan Ford (Ford *et al.*, 2014). For the computational analysis, we developed a pipeline called COVERnant that allows for easy analysis of ChIP-seq data (Wedel *et al.*, 2017).

3.3.2 Sites enriched in H2A.Z show increased sensitivity to MNase

To investigate whether an increased instability of H2A.Z-containing nucleosomes contributes to increased chromatin accessibility, I isolated total nucleosomal DNA from *T. brucei* using an H3 antiserum (Gassen *et al.*, 2012) in an MNase-ChIP-seq experiment and sequenced it on an Illumina HiSeq 2500 platform. A total of 18.6 million concordantly aligning 100 bp sequence reads could be mapped to the *T. brucei* 427v24 genome corresponding to an average genome coverage of ~53X.

As quality control we analyzed the fragment size distribution upon mapping of the sequencing reads and the rotational positioning of the nucleosomes by investigating dinucleotide frequencies within nucleosomal DNA (Figure 3.4). We plotted the frequency of all sequencing fragments in dependency of their size and found them to be distributed according to a Gaussian distribution that peaks at the standard size of nucleosomal DNA (147 bp, Figure 3.4A). Rotational nucleosome positioning is characterized by the DNA sequence associated with nucleosomes. Given that AT- and GC- dinucleotides differ in their rigidity, the formation of nucleosomes is dependent on the bending properties of the DNA. We extracted fragments of 147 bp in size and determined the rotational positioning of *T. brucei* nucleosome sequences. We found that the sequence of nucleosomal DNA is characterized by a 10 bp-periodicity of less rigid AA/AT/TA/TT dinucleotides (Figure 3.4B) as it has been previously shown for yeast (Brogaard *et al.*, 2012). This indicates, that dinucleotide patterns are important for the rotational positioning of trypanosome nucleosomes and both, the fragment size distribution and the rotational nucleosome positioning, suggest that my nucleosome occupancy maps are of high resolution.

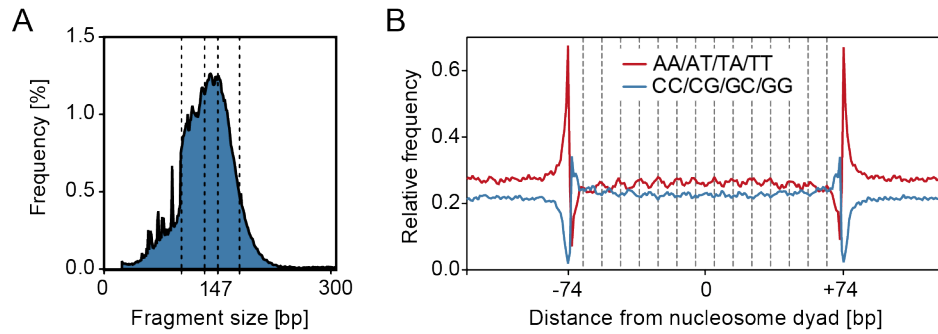


Figure 3.4 Fragment size distribution and dinucleotide frequencies upon MNase-ChIP-seq.

(A) Fragment size distribution after sequencing and joining of paired sequencing reads. Dashed lines indicate the fragment sizes 100, 137, 147 and 157 bp. (B) Relative frequencies of AA/AT/TA/TT and CC/CG/GC/GG dinucleotides throughout 147 bp of nucleosomal DNA for each bp relative to the nucleosome dyad. Dashed lines indicate distance of 10 bp from position -74 bp.

Given that MNase digestion of accessible chromatin results in shorter sequencing fragments than the nucleosome-specific 147 bp, I mapped sub-nucleosomal fragments with a size of 100-130 bp genome-wide and found sites enriched in H2A.Z to be enriched in those fragments (Figure 3.5A). The plotted H2A.Z distribution in Figure 3.5 to depict sites of H2A.Z enrichment at TSRs was obtained from an MNase-ChIP-seq experiment using a specific custom-made polyclonal H2A.Z antibody generated in this study (see chapter 2.4.2). When mapping supra-nucleosomal fragments with a size of >175 bp, I found sites of H2A.Z enrichment to be depleted of those fragments. To validate this observation, I plotted the averaged nucleosome occupancy across all 71 divergent and 57 non-divergent TSRs. This further supported an increase in sub-mono-nucleosomal DNA fragments across TSRs (Figure 3.5B). These findings indicate that DNA associated with H2A.Z-containing nucleosomes is more easily digested by MNase than DNA bound to nucleosomes containing canonical histones, revealing TSRs as regions of increased DNA accessibility.

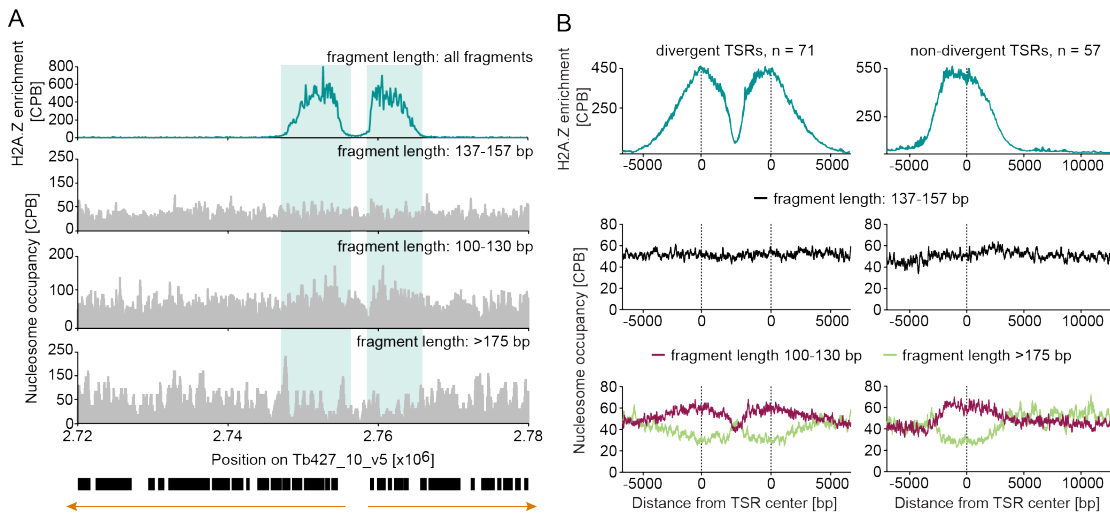


Figure 3.5 TSRs show increased MNase sensitivity.

(A) MNase-ChIP-seq data of H2A.Z-containing nucleosomes (cyan) and total nucleosomes (nucleosome occupancy, grey) grouped based on size of digestion products at a representative TSR on chr 10. Black boxes represent ORFs. Orange arrows indicate the direction of transcription. (B) The enrichment of H2A.Z and total nucleosome occupancy averaged across all divergent TSRs (left panel) and non-divergent TSRs (right panel) are plotted relative to the midpoint of the region between the TSRs, and the TSR center, respectively. Dashed lines mark the respective TSR centers.

3.4 Concluding remarks

Combining the mapping of the RNA pol II subunit RPB9 and short primary transcripts, I was able to demonstrate that RNA pol II-mediated transcription initiates directionally at the 5'-end of TSRs. Both, RPB9 and short primary transcripts, are enriched at the 5'-end of TSRs, where short primary transcripts are enriched within a region of ~2kb and the enrichment of RPB9 localizes ~200 bp downstream transcription initiation. Furthermore, by mapping nucleosomal DNA using MNase-ChIP-seq I could show that TSRs show an increased MNase sensitivity and thus are more accessible to proteins compared to the rest of the genome.

4

GT-rich DNA sequences can initiate transcription and induce targeted H2A.Z deposition

4.1	TSR DNA sequences are capable to initiate transcription	88
4.2	GT-rich promoter elements can trigger transcription initiation	93
4.3	GT-rich promoter elements promote targeted H2A.Z deposition	97
4.4	Concluding remarks	98

The localization of RNA pol II transcription initiation sites described in the previous chapter narrows the genome to putative promoter regions. Thus, it is now possible to analyze the DNA sequence within these regions and I aimed to investigate whether the DNA sequence is sufficient to initiate transcription. To this end, I chose sequences from candidate TSRs and inserted them with a downstream located reporter gene in a non-transcribed region in the *T. brucei* genome. If the inserted DNA sequence mediates transcriptional activity the gene product of the reporter gene can be measured. Furthermore, I investigated whether candidate DNA sequences are involved in the targeted deposition of the histone variant H2A.Z by conducting MNase-ChIP-seq with a custom-made polyclonal H2A.Z-specific antibody.

4.1 TSR DNA sequences are capable to initiate transcription

Even though much effort has been invested to identify RNA pol II promoter sequences for protein-coding genes in *T. brucei*, no promoter motifs have been described so far (Clayton, 2002). Thus, I aimed to either identify promoter sequences or to exclude a role of the DNA sequence during transcription initiation. To this end, I conducted a systematic approach to investigate the ability of DNA sequences to initiate transcription in *T. brucei in vivo* using a reporter assay. As potential candidate DNA sequences, I considered the regions identified in chapter 3 (Appendix Table 7.1) being enriched in primary transcripts, RNA pol II and the histone variant H2A.Z. I chose the candidate sequences based on three criteria: 1) The TSRs should be located on chromosome 10, since this is the most characterized so far. 2) I aimed to investigate the DNA sequence spanning the complete H2A.Z enrichment. Thus, I selected TSRs that did not contain genes spanning the H2A.Z enrichment to avoid partial translocations of genes, which may lead to secondary effects. 3) The selected TSRs should not contain NotI and XhoI restriction sites since those are required for the linearization of the final construct prior to transfection. Among 20 TSRs located on chromosome 10, two TSRs were found to meet the above-mentioned criteria (Figure 4.1). Both regions were chosen to be inserted along with a reporter gene in a transcriptional silent locus in the *T. brucei* genome to investigate whether the inserted DNA sequences can drive transcription initiation.

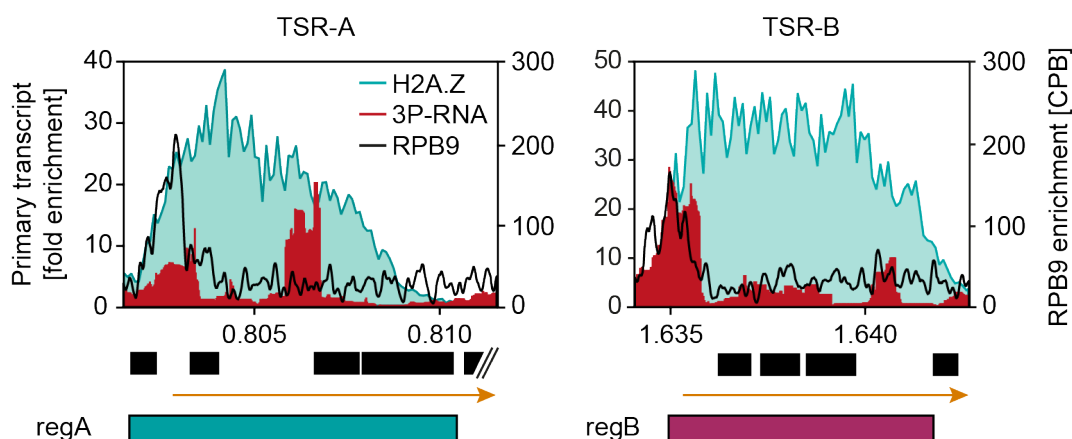


Figure 4.1 Examined TSR DNA sequences.

Primary transcript (red), RPB9 enrichment (black) and relative H2A.Z (cyan) distribution across the two tested TSRs TSR-A (left panel) and TSR-B (right panel) in dependency of the genomic position on chromosome 10. Black boxes indicate ORFs and orange arrows direction of transcription. The cyan bar represents the complete TSR-A DNA sequence regA (Tb427_10_v5:800,949-810,167; 9,218 bp) and the pink bar the complete TSR-B DNA sequence regB (Tb427_10_v5:1,634,960-1,641,653; 6,693 bp) examined in the reporter assay.

4.1.1 Insertion of TSR DNA sequences in transcriptional silent locus

I generated a construct, which allows the insertion of different DNA sequences upstream of a firefly luciferase gene (*FLUC*) into a transcriptional silent region. If the inserted DNA sequence mediates transcription initiation, *FLUC* will be transcribed and luciferase activity can be measured. Since the *T. brucei* genome is organized into PTUs most of the core genome is actively transcribed. Only short regions between PTUs have been described not to be transcribed (Kolev *et al.*, 2010; Siegel *et al.*, 2010), however those are found to be enriched in histone variants H3.V and H4.V, which are thought to repress transcription (Siegel *et al.*, 2009). Therefore, I decided to target the reporter construct between divergent TSRs on chromosome 1, since these regions contain low levels of H3.V and H4.V compared to other non-transcribed regions (Figure 4.2).

To generate the targeting construct, I used the backbone of pLEW100v5_HYG as it already contains a robust and well investigated rRNA promoter-driven *FLUC* gene containing the 5' UTR of a *GPEET* gene (including *trans*-splicing motifs), the *FLUC* CDS and the 3' UTR of aldolase (including a polyadenylation site). This construct was generated to inducibly overexpress a gene of interest, in this case *FLUC*, from a random rRNA locus. It contains a hygromycin phosphotransferase gene (*HYG*) that is transcribed by a T7 promoter divergent to *FLUC*. The targeting region within the construct consists of one sequence containing a NotI restriction site in the center. Upon NotI digest the construct is linearized and the complete construct is stably integrated into the genome via the homologous sequences at each end of

4. GT-rich DNA sequences can initiate transcription and induce targeted H2A.Z deposition

the linearized construct. In order to target this construct to another genomic location this system only requires one cloning step. In first attempts, in which I tried to replace the endogenous targeting sequence with a targeting sequence harboring a NotI restriction site in the center, as well, remained unsuccessful, since the integration into the *T. brucei* genome failed. Therefore, I replaced the targeting region downstream of *HYG* by an upstream homology region (UR) and inserted a downstream homology region (DR) downstream of *FLUC* generating the basal construct pCW24v2. To insert candidate DNA sequences in the non-transcribed locus the rRNA promoter is exchanged with the candidate sequence and integrated into the *T. brucei* genome. All constructs used in this study to target DNA sequences between the divergent TSR on chromosome 1 are derivatives of pCW24v2 and are listed in (Table 2.2).

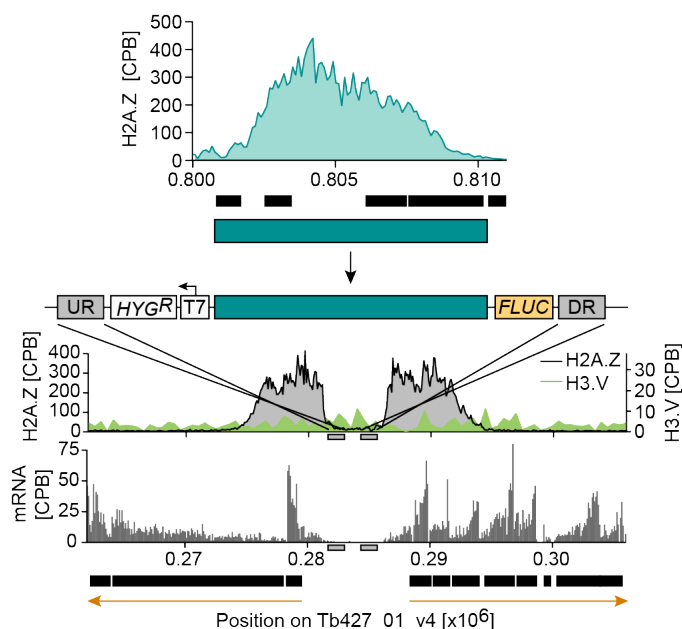


Figure 4.2 Approach to target TSR DNA sequences to a non-transcribed locus.

The TSR-A DNA sequence (cyan) is inserted into a targeting construct upstream of a firefly luciferase gene (*FLUC*). The construct is targeted to a non-transcribed locus between dTSRs on chromosome 1 (mRNA levels are shown in grey and were determined previously; Vasquez *et al.*, 2014) via upstream and downstream homology regions (UR, DR). H2A.Z and H3.V levels are shown in grey and green and were determined in this study and previously (Siegel *et al.*, 2009), respectively. Black boxes represent ORFs. Orange arrows indicate direction of transcription.

4.1.2 TSR DNA sequence-mediated transcription initiation is dependent on the genomic locus

To investigate the ability of the two TSR DNA sequences illustrated in Figure 4.1 to mediate transcription initiation, I generated two constructs based on pCW24v2 in which I replaced the rRNA promoter by either *regA* or *regB*. Additionally, I generated a 'no promoter control' in which

4. GT-rich DNA sequences can initiate transcription and induce targeted H2A.Z deposition

the rRNA promoter was removed. All four constructs were individually inserted into the region between two PTUs on chromosome 1 (Figure 4.3A, left panel) of SM cells expressing a renilla luciferase gene in the active expression site to account for differences in cell number. The insertion of the no promoter control resulted in no luciferase activity confirming that the target region is indeed not transcribed. Upon insertion of the rRNA promoter luciferase activity was detectable showing that the region is transcription permissive. The insertion of both constructs containing the TSR DNA sequences resulted in an 8.7-fold and 9.6-fold increase in luciferase activity compared to the no promoter control (Figure 4.3A, right panel). These results indicate that the DNA sequence found at TSRs contain elements able to initiate transcription.

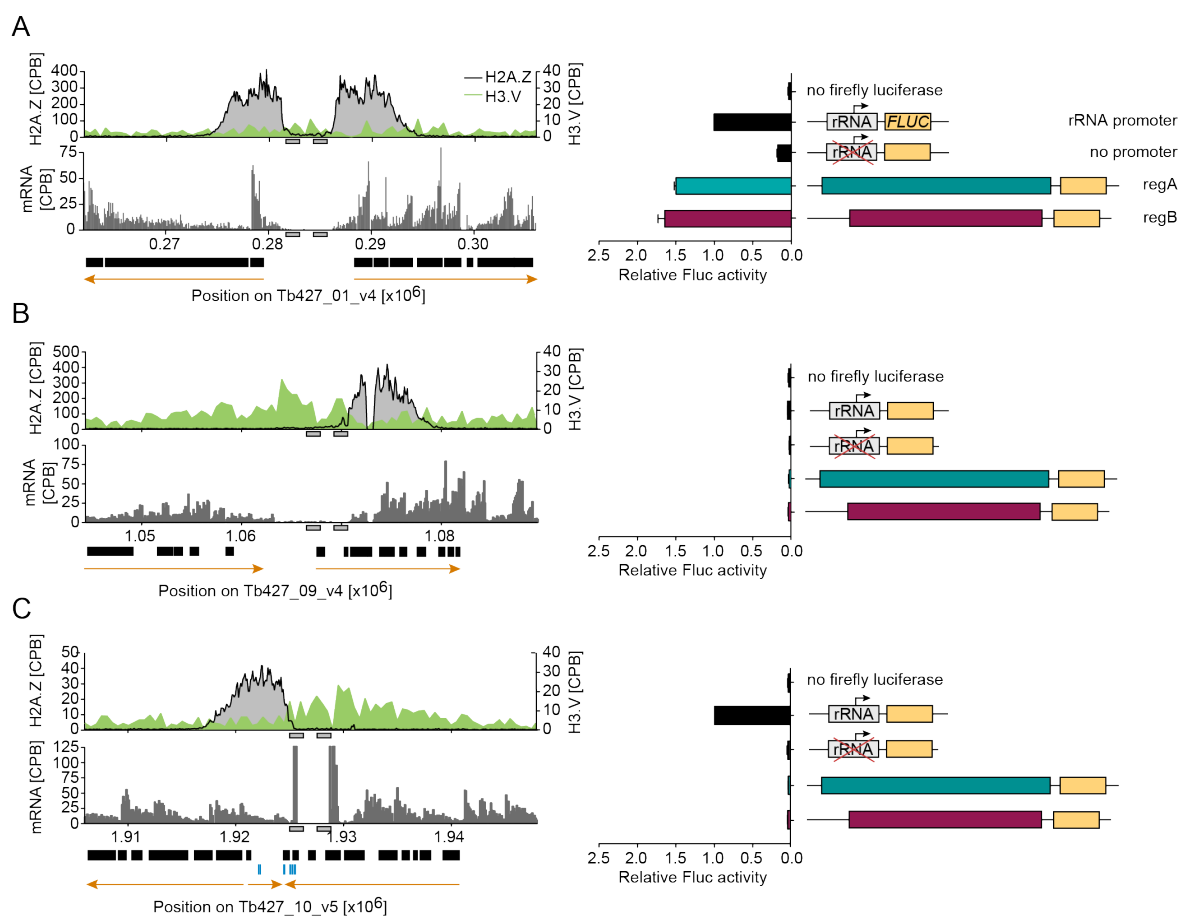


Figure 4.3 TSR DNA sequences are capable to mediate transcription initiation dependent on the genomic location.

(A) Targeting region of pCW24v2 (left panel) and luciferase activity upon insertion of regA and regB (right panel). (B) Targeting region of pCW27v2 (left panel) and luciferase activity upon insertion of regA and regB (right panel). (C) Targeting region of pCW28v2 (left panel) and luciferase activity upon insertion of regA and regB (right panel). H2A.Z, H3.V and RNA levels are shown in light grey, green and dark grey, respectively. Grey boxes represent regions of homology. Orange arrows indicate direction of transcription. To account for differences in cell number, Fluc activity was normalized to ectopically expressed Rluc activity. To account for technical variations, values were normalized to rRNA promoter-driven Fluc activity. Data are presented as mean \pm SD. Error bars indicate standard deviation between two replicates.

4. GT-rich DNA sequences can initiate transcription and induce targeted H2A.Z deposition

To investigate whether the tested DNA sequences are capable to initiate transcription at other transcriptional silent loci, I generated two additional targeting constructs, pCW27v2 and pCW28v2, to target a locus upstream of a ndTSR on chromosome 9 (Figure 4.3B, left panel) and a locus between tRNA genes transcribed by RNA pol III (Figure 4.3C, left panel), although they show increased levels of H3.V and H4.V. Targeting the control constructs to both locations revealed that both loci are transcriptional silent, whereas the rRNA promoter was able to induce transcription initiation only at the locus on chromosome 10 (Figure 4.3B and C, right panels). Targeting *regA* and *regB* to both loci did not result in luciferase expression underlining the importance of the genomic context for gene expression.

4.1.3 The transcription-mediating sequence element is distributed across TSRs and directs transcription

To identify specific DNA elements that are able to initiate transcription within the TSRs, I first investigated whether they are distributed across the TSR or concentrated in a specific region. I divided *regA* and *regB* in 7 and 5 evenly spaced fragments of 1800 bp in length with a 500 bp overlap between adjacent fragments. Inserting those 12 fragments in the locus targeted by pCW24v2 revealed that all 12 fragments were able to initiate transcription well above background (Figure 4.4). Fragments originating from the 5'-end of each TSR resulted in the highest luciferase activity (*regA1/2* and *regB1*) and luciferase activity tends to decrease towards the 3'-end. Interestingly, luciferase activity increases about 2-fold when inserting *regA5* compared to *regA4*. *RegA5* originates exactly downstream of an additional peak of primary transcripts. Taken together, these findings suggest a distribution of DNA sequence elements across the TSRs and consequently argue against the presence of well-defined canonical promoter motifs. Instead, the observed pattern is similar to that reported for dispersed promoters that lead to broad regions of transcription initiation (Deaton and Bird, 2011).

The analysis of primary transcripts in chapter 3.2 revealed that transcription initiates with a strong strand bias. To determine whether TSR-derived DNA sequences are able to direct transcription, I inserted the reverse complement sequence of the fragments that yielded the highest luciferase activity (Figure 4.4, *regA2rc*, *regB1rc*). Insertion of both sequences resulted in a decrease in luciferase activity by 4.7-fold and 3.3-fold, respectively. This is in good agreement with the 4:1 ratio of sense to antisense primary transcripts I observed and strongly suggests that transcription is directionally initiated. As a consequence, I hypothesize that the

responsible promoter elements are unevenly distributed across the coding and noncoding strand.

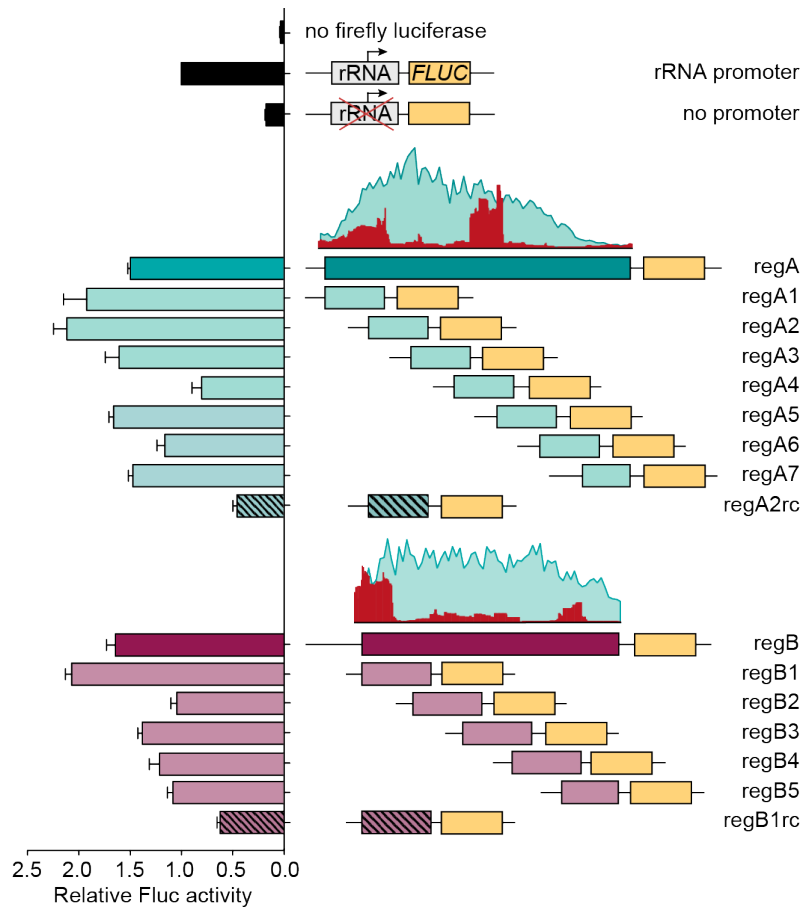


Figure 4.4 DNA elements are distributed across TSRs and provide directionality to transcription.

Luciferase assays performed after insertion of fragments derived from regA and regB. The fragments are 1800 bp in size and overlap with adjacent fragments by 500 bp. H2A.Z levels and primary transcript levels across regA and regB are shown in cyan and red, resp. Fragments, that were inserted as reverse complement are represented by striped bars. To account for differences in cell number, Fluc activity was normalized to rRNA promoter-driven Fluc activity. To account for technical variations, values were normalized to rRNA promoter-driven Fluc activity. Data are presented as mean \pm SD. Error bars indicate standard deviation between two replicates.

4.2 GT-rich promoter elements can trigger transcription initiation

To identify sequence elements being unevenly distributed between the coding and the noncoding strand across TSRs, we computationally divided 199 TSRs into 5 evenly spaced regions according to their H2A.Z enrichment and in each region, we searched for sequences of 10 bp in size (10mers) that are at least 6-fold enriched on the coding strand compared to the noncoding strand. The vast majority of enriched 10mers contained either long stretches of

Gs or Ts or were rich in Gs and Ts in general and we found most of them to be enriched at the 5'-end of TSRs (Figure 4.5A and B, Appendix Table 7.2).

To test whether these GT-rich sequences are able to mediate directional transcription, I designed two synthetic GT-rich promoter sequences in which I assembled the enriched 10mer sequences. The sequences were designed in a way that they contained as many GT-rich 10mers as possible and that they still met the synthesis requirements of Integrated DNA Technologies (IDT). Thus, where necessary, As and Cs were added to reduce the GT content and to allow synthesis. As a result, a short 210 nt promoter sequence with a GT content of 80% and a longer 416 nt promoter sequence with a GT content of 81% after the fusion of two short fragments could be obtained (Figure 4.5C). The promoter sequences were inserted into the pCW24v2 targeting construct and integration of the two constructs resulted for both in high luciferase activity of 21-fold and 17-fold higher compared to the no promoter control (Figure 4.5D). In addition, direction of transcription was highly dependent on the GT-rich sequence demonstrated by the strong decrease of luciferase activity upon insertion of the respective reverse complement sequence of the GT-rich promoter sequences. The luciferase activity mediated by the GT-rich promoter sequences is 2.4-fold and 2.0-fold higher compared to the luciferase activity measured upon insertion of endogenous TSR sequences (compare Figure 4.3A, right panel and Figure 4.5D). To test whether the transcriptional activity of regA and regB was too low to induce transcription initiation in the genomic locations with increased H3.V and H4.V levels (Figure 4.3B and C, left panels), I inserted GT_416_nt instead. Even in a cell line lacking H3.V no luciferase activity mediated by the stronger GT-rich promoter sequence could be detected (Figure 4.5E), suggesting the involvement of additional factors, such as the histone variant H4.V and the genomic location.

4. GT-rich DNA sequences can initiate transcription and induce targeted H2A.Z deposition

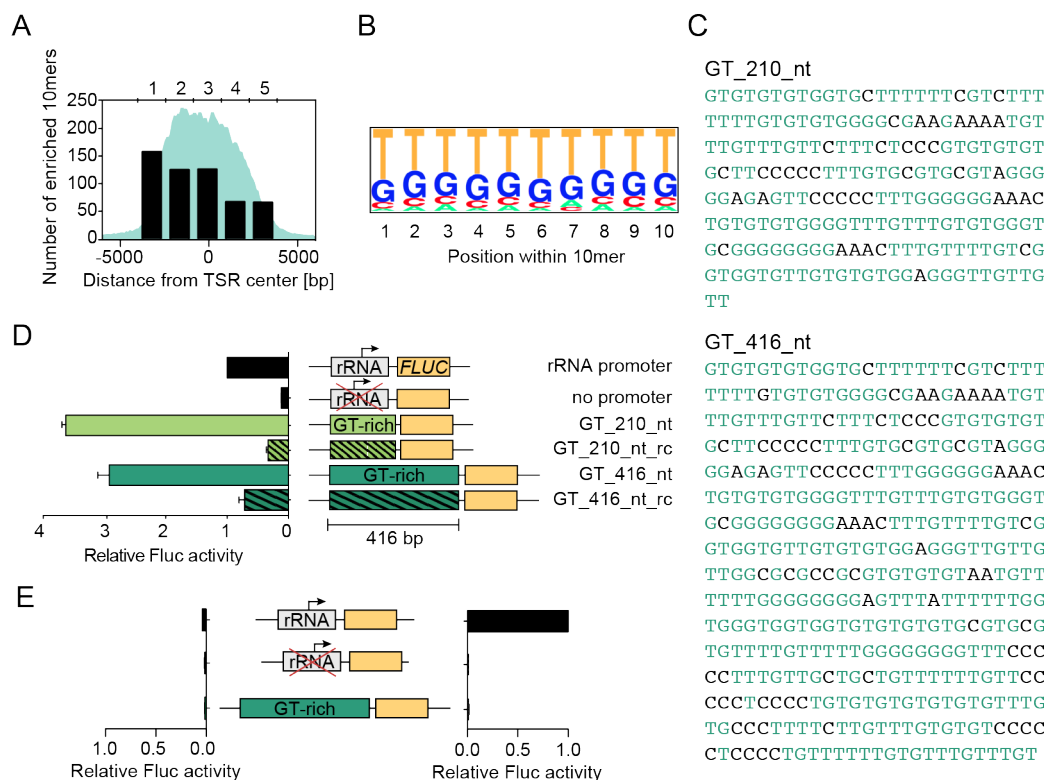


Figure 4.5 GT-rich sequence elements on the coding strand mediate transcription initiation.

(A) All TSR sequences were divided into 5 equal regions and the number of 10mers enriched at least 6-fold on the coding strand compared to the noncoding strand were counted. The mean H2A.Z enrichment is shown in cyan. (B) The consensus sequence of 10mers identified in (A) within region 1 was calculated using pictogram (Burge *et al.*). (C) Sequence of synthetic GT-rich promoters composed of the most enriched 10mers. (D) Luciferase assays after insertion of the two synthetic GT-rich promoter sequences GT_210_nt (light green) and GT_416_nt (dark green) and their respective reverse complement sequence (striped bars). (E) Luciferase assays after insertion of the synthetic GT-rich promoter sequence GT_416_nt into the genomic locations targeted by pCW27v3 (left panel) and pCW28v3 (right panel) in BFJEL25. To account for differences in cell number, Fluc activity was normalized to ectopically expressed Rluc activity. To account for technical variations, values were normalized to rRNA promoter-driven Fluc activity. Data are presented as mean \pm SD. Error bars indicate standard deviation between two replicates.

To compare the transcriptional activity between two dTSRs on chromosome 1 induced either by the GT-rich sequence GT_210_nt or the rRNA promoter to endogenous RNA pol II and RNA pol I transcription levels, I targeted a promoter-less *FLUC* within a PTU and a rRNA promoter driven *FLUC* to the rRNA spacer region. Insertion of the luciferase gene within a PTU resulted in a 11.6-fold higher luciferase activity compared to the luciferase activity mediated by GT_210_nt. After insertion of the rRNA promoter driven luciferase into a rRNA spacer region, I find the luciferase activity to be 38.7-193.6-fold higher compared to the region between dTSRs (Figure 4.6B), which can be explained by the compartmentalization of the nucleus, since the vast majority of RNA pol I is spatially restricted to the nucleolus. In addition, I find luciferase activity to vary greatly among different clones, which is probably due to the integration in different rRNA spacer regions already described previously (Alsford *et al.*, 2005).

4. GT-rich DNA sequences can initiate transcription and induce targeted H2A.Z deposition

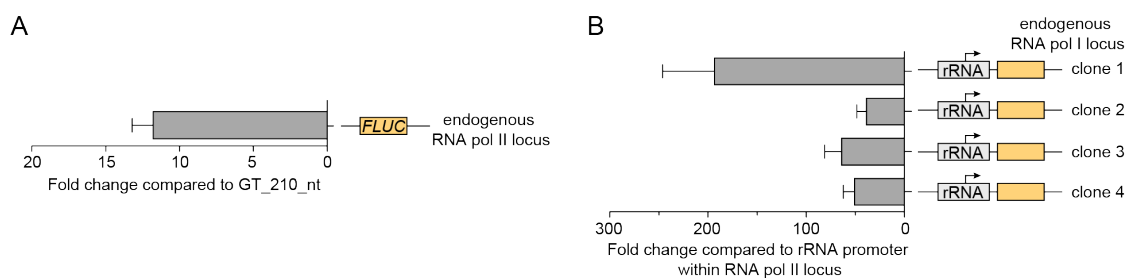


Figure 4.6 Comparison of endogenous expression levels.

(A) To compare GT_210_nt-mediated luciferase activity to endogenous RNA pol II transcription levels a promoter-less *FLUC* has been inserted in a PTU on chromosome 1 (Tb427_01_v4:500,640-501,239) using the construct pCW37. (B) To compare rRNA promoter driven luciferase activity between dTSRs to endogenous RNA pol I levels rRNA promoter driven *FLUC* was inserted in a rRNA spacer using the construct pLEW100v5_HYG. Luciferase activity was measured of 4 different clones. To account for differences in cell number, Fluc activity was normalized to ectopically expressed Rluc activity. To account for technical variations, values were normalized to rRNA promoter-driven Fluc activity. Data are presented as mean \pm SD. Error bars indicate standard deviation between two replicates.

These results indicate that the newly identified GT-rich sequence elements are able to induce directional transcription initiation. However, the resulting transcriptional activity is lower than those resulting from endogenous RNA pol II transcription initiation.

To investigate whether this additional site of transcription initiation affects the transcription of flanking PTUs, I performed a qPCR analysis to compare transcript levels of different genes located upstream and downstream of the insertion site upon insertion of GT_416_nt. The genes were chosen based on the wild type expression level and the distance from the insertion site. I analyzed the expression of Tb427.01.860, Tb427.01.890, Tb427.01.990 and Tb427.01.1050 in the no promoter control and two clones of SM Rluc GT_416_nt, whereas Tb427.01.890 and Tb427.01.990 are the closest genes with a distance to the insertion site of 5 kb, each and Tb427.01.860 and Tb427.01.1050 are 22 kb and 17 kb apart, respectively. The insertion of GT_416_nt had no large effect on the transcript levels of genes located downstream of the insertion site. Upstream located genes were slightly negatively affected, possibly by run-through transcription of the T7 polymerase transcribing the resistance marker in the targeting construct (Figure 4.7).

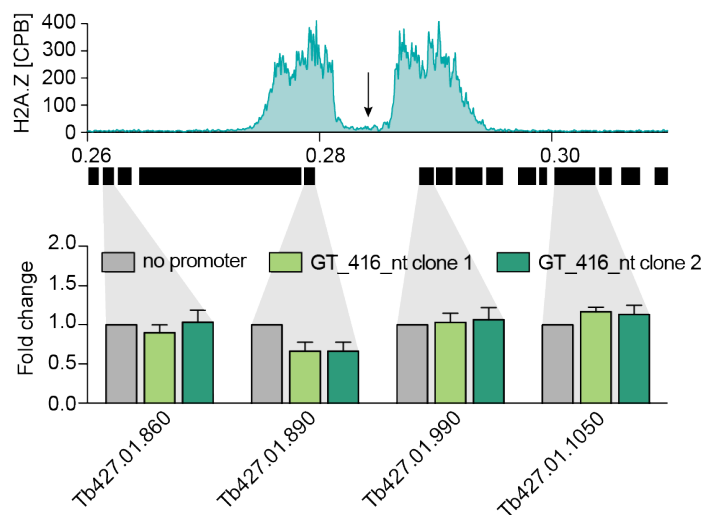


Figure 4.7 Influence of GT-rich sequence insertion on the transcription of flanking PTUs.

Comparison of transcript levels of genes located in PTUs flanking the insertion site (black arrow) before and after insertion of GT_416_nt. Transcript levels of genes in the no promoter control are shown in grey and are set to 1. Transcript levels of genes after the insertion of GT_416_nt compared to the no promoter control are shown for two clones in light and dark green. Data are presented as mean \pm SD. Error bars indicate standard deviation among triplicates.

4.3 GT-rich promoter elements promote targeted H2A.Z deposition

The incorporation of H2A.Z into promoter-nucleosomes is the result of a cascade of protein binding and histone modification events, which is induced by the binding of RNA pol II to the promoter and phosphorylation within its CTD. One of the final steps is the acetylation of histones H4 and H3. The acetylation marks recruit the bromodomain-containing chromatin remodeling complex SWR, which replaces H2A by H2A.Z. Since the in chapter 4.2 identified GT-rich sequences were capable to induce transcription initiation, I aimed to investigate whether they also contribute to the H2A.Z deposition to the site of insertion.

Thus, I performed an MNase-ChIP-seq experiment using the cell lines SM Rluc GT_210_nt and SM Rluc GT_416_nt and a custom-made polyclonal H2A.Z antibody (see chapter 2.4.2 for information about its generation). The mapping of the genome-wide distribution of H2A.Z in these cell lines revealed that both loci contained H2A.Z-containing nucleosomes. However, H2A.Z levels across both GT-rich sequences were lower compared to an endogenous TSR (Figure 4.8A). Since the formation of chromatin structure is a dynamic process, it is possible that the level of H2A.Z enrichment and transcriptional activity may increase with the number of cell divisions. Therefore, I re-generated the cell lines SM Rluc GT_210_nt and SM Rluc GT_416_nt and performed luciferase assays 8 and 30 days post transfection. The measurements revealed an increase in luciferase activity over time (Figure 4.8B), supporting the hypothesis that H2A.Z recruitment may increase over time, as well. Thus,

4. GT-rich DNA sequences can initiate transcription and induce targeted H2A.Z deposition

I additionally performed an MNase-ChIP-seq experiment 8 days post transfection. For the short GT-rich sequence I indeed observed a time-dependent increase in H2A.Z levels, However, for the long GT-rich sequence no such increase was detectable.

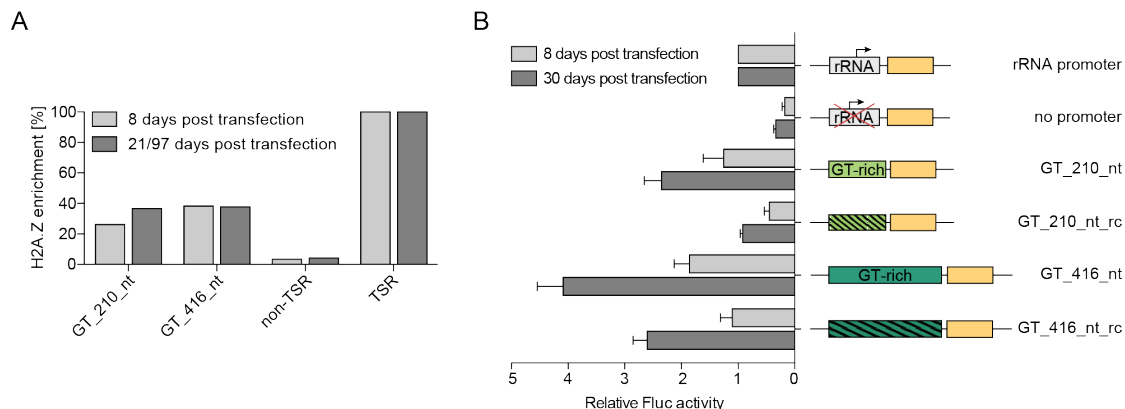


Figure 4.8 H2A.Z enrichment and luciferase activity increase over time.

(A) H2A.Z enrichment across GT_210_nt, GT_416_nt, a 6 kb region upstream of the adjacent TSR (non-TSR) and the TSR upstream of the site of insertion. The H2A.Z levels were determined by MNase-ChIP-seq for SM Rluc GT_210_nt and SM Rluc GT_416_nt 8 and 21 days and 8 and 97 days post transfection, respectively. The H2A.Z enrichment of the adjacent TSR was set to 100%. (B) Luciferase activity was measured 8 and 30 days post transfection. To account for differences in cell number, Fluc activity was normalized to ectopically expressed Rluc activity. To account for technical variations, values were normalized to rRNA promoter-driven Fluc activity. Data are presented as mean \pm SD. Error bars indicate standard deviation between two replicates.

4.4 Concluding remarks

The insertion of different complete or partial TSR DNA sequences into non-transcribed loci demonstrated that directed transcription initiation in *T. brucei* is indeed DNA-sequence-mediated. However, the ability of the tested DNA sequences to initiate transcription is highly dependent on the genomic context. Analysis of TSR DNA sequences revealed GT-rich sequences enriched on the coding strand to be the key element to trigger directed transcription initiation. In addition, the GT-rich sequences promoted the targeted H2A.Z deposition to the site of the inserted GT-rich promoters.

5

Nucleosome depleted regions at exon boundaries are affected by the DNA sequence

5.1	Exon boundaries rather than TSRs contain well-defined NDRs	100
5.2	Composition of the polyY tract affects gene expression and nucleosome positioning..	102
5.3	Concluding remarks	104

In a wide range of organisms dispersed promoters have been identified (Martínez-Calvillo *et al.*, 2003; Zhang and Dietrich, 2005; Saxonov *et al.*, 2006; Yamamoto *et al.*, 2009; Kolev *et al.*, 2010; Ni *et al.*, 2010; van Heeringen *et al.*, 2011). Those are characterized by several transcription initiation sites within regions of 50 bp-10 kb and a lack of defined promoter motifs (Carninci *et al.*, 2006; Sandelin *et al.*, 2007; Koch *et al.*, 2011). Furthermore, they show a higher enrichment in H2A.Z and harbor broader NDRs compared to focused promoters (Tirosh and Barkai, 2008; Rach *et al.*, 2011). In addition to promoters, exon/intron boundaries have been shown to be depleted from nucleosomes while exons are highly occupied by nucleosomes compared to introns (Schwartz *et al.*, 2009; Tilgner *et al.*, 2009; Chen *et al.*, 2010). It has been proposed that this increase slows down the transcription rate facilitating the co-transcriptional recruitment of splicing factors (Naftelberg *et al.*, 2015).

The data presented in the previous two chapters provided evidence, that TSRs in *T. brucei* could serve as dispersed promoters. They are enriched in H2A.Z over regions of 7-10 kb, which appear to have a more open chromatin structure and RNA pol II-transcription initiates within ~2 kb at the 5'-end of H2A.Z enrichment. Although transcription initiation is DNA-sequence-mediated, defined promoter motifs are missing. In this chapter, I investigated whether TSRs and splice sites in *T. brucei* contain NDRs by mapping nucleosome positioning.

5.1 Exon boundaries rather than TSRs contain well-defined NDRs

To investigate whether TSRs contain NDRs I performed an MNase-ChIP-seq experiment using a specific H3 antiserum (Gassen *et al.*, 2012) to pull down nucleosomes from TSRs and located within PTUs. The digestion of the chromatin with MNase only leaves DNA that was associated with the histone octamer allowing a precise localization of nucleosomes genome-wide when the nucleosomal DNA is sequenced and mapped back to the genome. Depending on how many reads align to a certain region of the genome, it can be determined whether a region is highly occupied by nucleosomes or depleted from nucleosomes. To analyze H3 MNase-seq data, Konrad Förstner developed the tool 'COVERnant' that calculates the nucleosome occupancy from the number of aligned reads and normalizes to the input control and the total number of reads (Wedel *et al.*, 2017). In addition, it allows the generation of metaplots, in which the nucleosome occupancy can be averaged for selected genomic regions.

Given the broad regions of dispersed transcription initiation and the organization of genes into PTUs I analyzed the nucleosome occupancy across the first gene of each PTU. I used the ATG of the first genes as a proxy for transcription initiation and averaged the data for the 184 genes. The metaplots revealed a strong depletion of H3-containing nucleosomes

~90 bp upstream of the ATG (Figure 5.1A and B). However, my primary transcript data indicated that RNA pol II transcription initiates much further upstream of the ATG of the first gene of a PTU (Figure 4.1, TSR-B). In addition, due to the median 5'UTR length of ~90 bp in *T. brucei* (excluding the spliced-leader RNA; Siegel *et al.*, 2011) the here observed NDR might co-localize with the 5'-end of 5'UTRs, which contains an important motif for splicing. Thus, these data point to a role of NDRs in RNA processing rather than transcription initiation.

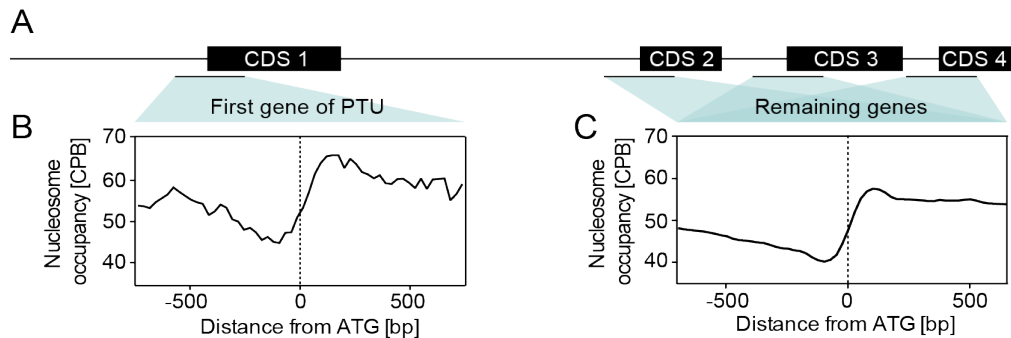


Figure 5.1 Most genes within PTUs are preceded by an NDR.

(A) Schematic illustration of PTU. (B) Average nucleosome occupancy plotted relative to the ATG of the first genes within PTUs ($n = 184$). The first genes have been defined according to a previous study (Kolev *et al.*, 2010) and the coordinates have been adjusted to the genome version Tb927v24. (C) Average nucleosome occupancy plotted relative to the ATG of all genes except of the first genes within PTUs ($n = 12,220$).

Given that *T. brucei* splices its polycistronic pre-RNA in *trans*, the 5'-end of each 5'UTR serves as a splice acceptor site (SAS). Thus, should exon boundaries be depleted of nucleosomes, I would expect to find an NDR upstream of each ATG. To test this hypothesis, I averaged the nucleosome occupancy for all remaining genes of the *T. brucei* genome, i.e. all genes excluding the first genes of PTUs, and observed a similar pattern as depicted in Figure 5.1B, indicating that most genes are preceded by an NDR (Figure 5.1C). Additionally, both metaplots show that the gene bodies located downstream of the NDR are more occupied by nucleosomes than regions upstream of the NDR.

It is proposed that nucleosomes act as 'speed bumps' and thereby slow down the rate of RNA pol II elongation and enhance co-transcriptional splicing efficiency (Nafteberg *et al.*, 2015). If this is the case, I would expect that RNA pol II levels correlate with nucleosome occupancy. Assuming that efficient *trans*-splicing results in high RNA levels, I would expect the amount of RNA to correlate with nucleosome occupancy. Thus, I grouped all genes of *T. brucei* based on the level of detected transcripts (Vasquez *et al.*, 2014) in those, which are highly (top 25 %), intermediately (middle 25 %) and weakly expressed (bottom 25 %). I generated metaplots, in which I averaged my RNA pol II-ChIP and H3 MNase-ChIP-seq data

across all genes assigned to the individual groups and plotted them to their ATG or SAS. Since the annotation of UTRs has been performed only for highly expressed genes so far, I used for those genes the coordinates of the SAS and for the remaining the coordinates of the ATG to avoid introducing a bias in RNA levels during the grouping. The analysis shows, that the RNA pol II coverage closely resembles those of the nucleosome occupancy and I find a well-defined NDR upstream of highly expressed genes, while it is absent upstream of genes that show low RNA levels (Figure 5.2). Notably, the NDRs of genes with intermediate and high RNA levels coincide with polyY tracts upstream of the genes analyzed. Plotting the data to ATGs exclusively shows the same results (Appendix Figure 7.2). PolyY tracts are short pyrimidine-rich sequences that are located upstream of the SAS and serve as binding sites for the U2AF65 subunit of the spliceosome (Kielkopf *et al.*, 2001). Since polyY tracts are primarily composed of stretches of thymines we determined their location by screening for 10mers of Ts upstream of the SAS while an interruption of one non-T base was allowed (Figure 5.2, lower panels). Taken together, my nucleosome positioning data suggest that NDRs are involved in the regulation of RNA processing rather than of transcription initiation.

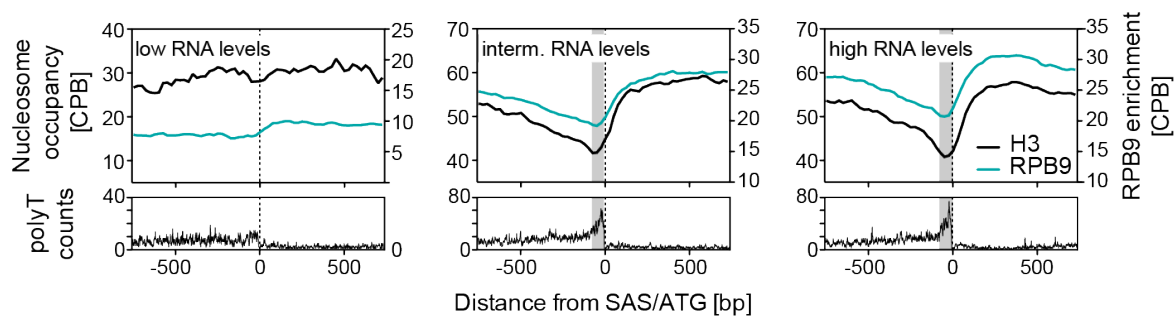


Figure 5.2 Nucleosome depletion correlates with the level of gene expression.

Average nucleosome occupancy (black) and RPB9 enrichment (cyan) plotted relative to the SAS/ATG of the 25 % of genes yielding the highest RNA levels (upper left panel, $n = 2,753$), the 25 % of genes yielding intermediate RNA levels (upper middle panel, $n = 2,753$) and the 25 % of genes yielding the lowest RNA levels (upper right panel, $n = 2,753$). RNA levels were determined previously (Vasquez *et al.*, 2014). The number of polyT tracts composed of 10mers of Ts (Hamming distance = 1) within the regions analyzed in the upper panels was counted and plotted relative to the SAS/ATG (lower panels). The polyT enrichment is highlighted in grey.

5.2 Composition of the polyY tract affects gene expression and nucleosome positioning

To investigate, how the formation of an NDR at exon boundaries is regulated, I investigated whether the DNA sequence of the T-rich polyY tract is involved in this process. On the DNA level homopolymeric sequences in general are intrinsically rigid and are disfavored during nucleosome formation (Suter *et al.*, 2000). Previous studies in which *T. brucei* cells were

transiently transfected with plasmids carrying *FLUC* preceded by polyY tracts of distinct composition showed that length and composition of polyY tracts influence *trans*-splicing (Siegel *et al.*, 2005). Since the effect of nucleosome positioning could not be evaluated in these experiments, I conducted a similar approach based on a stable integration of the reporter construct into the genome.

I generated two transgenic cell lines similar to SM Rluc GT_210_nt, which is described in chapter 4.2 and contains the short GT-rich promoter and *FLUC* preceded with the endogenous GPEET polyY tract. In the first cell line *FLUC* is preceded with a long T-rich polyY tract that has previously been shown to mediate highly efficient *trans*-splicing (Siegel *et al.*, 2005). In the third cell line no polyY tract is present upstream of *FLUC*. Luciferase assays revealed the highest luciferase activity for the GPEET polyY tract. The activity for the long T-rich polyY tract was twofold lower compared to the GPEET polyY tract and no luciferase activity was measured for the cell line without polyY tract (Figure 5.3). These measurements are in good agreement with the previous transient transfection experiments and underline the importance of the polyY tract for efficient *trans*-splicing.



Figure 5.3 Composition of polyY tract affects gene expression.

Luciferase assays after insertion of a *FLUC* reporter construct containing the short GT-rich promoter. *FLUC* was preceded either with the endogenous GPEET polyY tract (green), a long T-rich polyY tract (light blue) or no polyY tract (dark blue). Pyrimidines are shown in blue. To account for differences in cell number, Fluc activity was normalized to ectopically expressed Rluc activity. To account for technical variations, values were normalized to rRNA promoter-driven Fluc activity. Data are presented as mean \pm SD. Error bars indicate standard deviation between two replicates.

To evaluate the importance of the polyY tract for nucleosome positioning, I performed an MNase-ChIP-seq experiment against histone H3 to examine the nucleosome positioning around the 5'-end of the 5'UTR of *FLUC* within the three cell lines. The paired-end sequenced reads were joined to fragments, mapped to the genome and high-resolution nucleosome occupancy maps were generated using 'NUCwave' (Quintales *et al.*, 2015). 'NUCwave' calculates the position of a nucleosome according to the midpoint of the paired-end fragment. In order to obtain a higher resolution of the nucleosome center the fragments are symmetrically trimmed.

Analyzing the nucleosome occupancy maps revealed an extension of the NDR in the cells harboring the long T-rich polyY tract (Figure 5.4, middle panel) compared to those with the endogenous polyY tract (Figure 5.4, upper panel) and those without polyY tract (Figure 5.4, lower panel). This might be the consequence of the homopolymeric nature of the long T-rich sequence, which inhibits nucleosome formation. Additionally, these data indicate the presence of an NDR upstream of *FLUC* in the absence of a polyY tract (Figure 5.4, lower panel). Comparing the nucleosome positioning across the *FLUC* OFR it becomes apparent, that in all three cell lines *FLUC* is highly occupied by nucleosomes. However, in those in which *FLUC* is expressed the pattern is similar to each other (Figure 5.4, upper and middle panel) and different compared to the non-expressed *FLUC* (Figure 5.4, lower panel). In addition, the pattern suggests that in all three cell lines the GT-rich promoter is strongly depleted of nucleosomes, which might be due to the long homopolymeric stretches of Gs and Ts present in the promoter element. Taken together my results indicate that the polyY tract is important for efficient *trans*-splicing and can affect nucleosome positioning. However, there need to be additional elements that are involved in the generation of NDRs at exon boundaries.

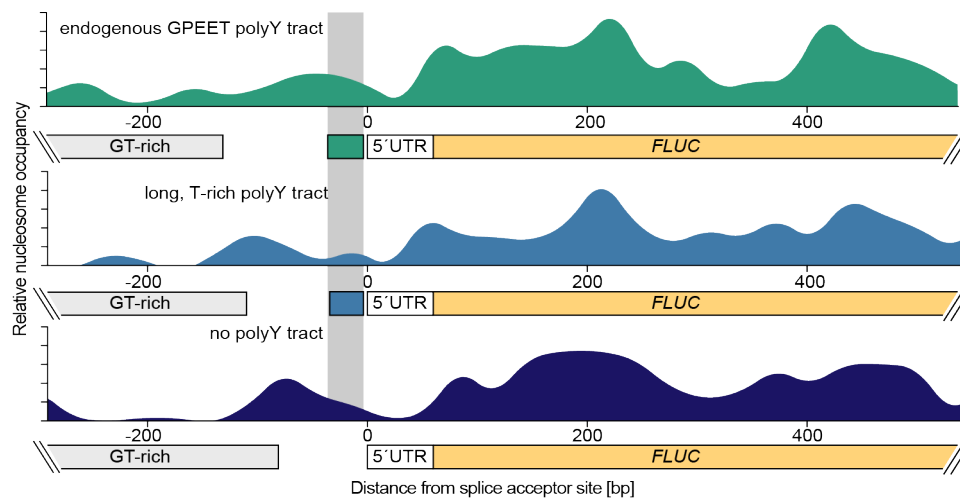


Figure 5.4 Composition of polyY tract affects nucleosome positioning.

Nucleosome occupancy was determined for the three cell lines described in Figure 5.3 by and aligned to the SAS of *FLUC*. The maps were generated from histone H3 MNase-ChIP-seq data processed with bowtie 1.1.1 and default NUCwave settings (Quintales *et al.*, 2015). The location of the respective polyY tract is highlighted in grey.

5.3 Concluding remarks

While the data generated within this study provide evidence that well-defined NDRs are absent from TSRs, they show that the 5'UTR of each gene is preceded by an NDR. Given the organization of genes in *T. brucei* in PTUs, these findings point to a role of NDRs in RNA

processing rather than transcription initiation. Particularly, when investigating the nucleosome occupancy of genes yielding high, intermediate and low RNA levels it became apparent that the presence of NDRs correlates with the level of gene expression. The analysis of the DNA sequence around the location of NDRs revealed that NDRs coincide with polyY tracts that are crucial for efficient *trans*-splicing. Manipulation of the DNA sequence of the polyY tract showed that its composition affects nucleosome depletion but is not the sole contributor in NDR formation. Additionally, I was able to show that RNA pol II enrichment correlates with nucleosome occupancy supporting the hypothesis that nucleosomes within exons, that are generally highly occupied by nucleosomes, act as 'speed bumps' (Naftelberg *et al.*, 2015). Taken together, these data highly suggest the chromatin structure as an important factor during post-transcriptional regulation of gene expression.

6

Discussion

6.1	Transcription initiates at the 5'-end of TSRs	107
6.2	RNA pol II transcription initiation is DNA sequence-mediated.....	108
6.3	GT-rich promoter elements contribute to targeted H2A.Z deposition	111
6.4	NDRs regulate gene expression post-transcriptional and are affected by the DNA sequence.....	113
6.5	Conclusion.....	114

6.1 Transcription initiates at the 5'-end of TSRs

Genome-wide studies performed in different eukaryotes revealed a strong association of constitutively expressed genes with dispersed promoters, which lack well-defined sequence motifs. *T. brucei*, an evolutionarily highly divergent eukaryote, lacks transcriptional regulation of RNA pol II-transcribed genes and promoter sequences have remained elusive. Thus, how RNA pol II-mediated transcription initiation is facilitated in this parasite remains a fundamental question. From studies in different eukaryotes it has been shown that transcription initiation sites can be identified by mapping primary transcripts and that RNA pol II-mediated transcription initiation sites are characterized by the enrichment of RNA pol II due to promoter proximal pausing. Additionally, it has been shown that specific PTMs and histone variants localize to sites of transcription initiation and that those are involved in establishing an open chromatin structure (see chapter 1.2 for references). In *T. brucei*, it has been shown that the 5'-ends of PTUs (TSRs) are enriched in PTMs, histone variants H2A.Z and H2B.V and that primary transcripts originate from these locations (see chapter 1.4 for references). However, where RNA pol II-mediated transcription exactly initiates remained to be elaborated.

The findings in this work demonstrate that RNA pol II-specific transcription initiates at the 5'-end of each PTU. Furthermore, they show that the chromatin structure is more open within TSRs compared to non-TSR regions. The co-localization of both, RNA pol II and primary transcripts, shows that RNA pol II-specific transcription initiation starts at the 5'-end of each TSR. The genome-wide localization of RNA pol II and primary transcripts not only enabled the localization of RNA pol II-mediated transcription initiation. It also allowed us to draw conclusions about the nature of transcription initiation in *T. brucei*. The enrichment of primary transcripts within regions of ~2 kb suggests that transcription initiates dispersedly within these regions. *T. brucei* lacks transcriptional control and thus all genes arranged in PTUs are constitutively transcribed. Thus, these results are in good agreement with findings in other organisms, in which transcription of constitutively expressed genes has been linked to dispersed promoters. In addition, the primary transcript data provide evidence that transcription initiation is directional. Furthermore, in other organisms it has been shown that RNA pol II enrichment is due to promoter-proximal pausing. In *T. brucei*, RNA pol II enrichment localizes 100-200 bp downstream of transcription initiation, which is in good agreement with findings in other organisms and suggests that *T. brucei* utilizes similar mechanisms during transcription as metazoans. This is surprising, since the predominant usage of polycistronic transcription in *T. brucei* is more common to prokaryotic systems. Given its early branching from the eukaryotic lineage in evolution, *T. brucei* could utilize mechanisms similar to yeast. Yeast, however, lacks the mechanism of promoter proximal pausing (Adelman and Lis, 2012). The

compaction of the chromatin influences the accessibility of DNA to proteins and thus, transcriptional activity. The chromatin state can be influenced by e.g. the incorporation of histone variants into nucleosomes. H2A.Z-containing nucleosomes have been shown to be less stable. However, unlike other eukaryotes, in which only single nucleosomes flanking the transcription initiation site contain H2A.Z, H2A.Z-containing nucleosomes are enriched within TSRs that span regions of ~10 kb. Results obtained from this study demonstrate that TSRs are more accessible to MNase digest compared to the remnant genome. These findings suggest that chromatin within TSRs is less compact than the remnant genome and agree with the findings obtained in other eukaryotes. However, there are substantial differences regarding the size of open chromatin regions. While regions associated with open chromatin in yeast measure 159 bp in mean (Lee *et al.*, 2013), the observed open chromatin regions at TSRs are ~10 kb in size. Given that data in this work provide evidence that NDRs are absent from TSRs, the observed broad regions of open chromatin reflect the absence of transcriptional regulation.

Taken together, the findings provide evidence for the conservation of the association between dispersed transcription initiation and constitutively expressed genes.

6.2 RNA pol II transcription initiation is DNA sequence-mediated

Dispersed transcription initiation has been shown to be associated with the lack of well-defined promoter sequences. In this work, I sought to elucidate whether transcription initiation at TSRs is DNA sequence-mediated and thus to shed light on the mechanism of RNA pol II-mediated transcription initiation in *T. brucei*.

Results from this work demonstrate that transcription is DNA sequence-mediated and GT-rich promoter elements have been identified to drive directional transcription initiation. Initial findings here show that the DNA sequence within endogenous TSRs is sufficient to initiate transcription. This has been demonstrated for the DNA sequence of both, complete TSRs and fragments of those. Fragments derived from the 5'-end of TSRs yielded the highest transcriptional activity, which decreased towards the 3'-end. The results obtained in this study are the first demonstration showing that the DNA sequence stably integrated into the genome mediates transcription initiation. The finding that all TSR fragments tested promote transcription initiation suggests the absence of well-defined promoter motifs and further supports the hypothesis that transcription is dispersedly initiated in *T. brucei*. The highest luciferase activity observed at the 5'-end of TSRs correlates with the results obtained from the sequencing of primary transcripts. A fragment in the center of TSR-A yielding again high

luciferase activity and co-localizing to a second peak of primary transcripts within the TSR reflects this correlation (Figure 4.1).

However, the ability of specific DNA sequence elements to initiate transcription is dependent on the genomic location. Insertion of a TSR sequence between two divergent TSRs resulted in transcriptional activity, whereas insertion of the same construct into a region enriched in H3.V led to no activity. The insertion of GT-rich promoter elements resulted in the same site-specific effects. In addition, insertion of an rRNA promoter between two divergent TSRs yielded lower transcriptional activity compared to its insertion into an rRNA array. These findings may be explained by the nuclear architecture of the *T. brucei* genome. Several previous studies revealed a clustering of specific RNA pol I, RNA pol II and RNA pol III subunits in distinct loci within the nucleus of *T. brucei* (Navarro and Gull, 2001; Uzureau *et al.*, 2008; Alsford and Horn, 2011). Most strikingly, these analyses showed that RNA pol I is restricted to the nucleolus and the expression site body, where the variant surface antigens are transcribed (Navarro and Gull, 2001). Another study reported that RNA pol I-driven transcriptional activity is 20-fold higher compared to those of RNA pol II (Wirtz and Clayton, 1995; Biebinger *et al.*, 1996). This may explain the differences in rRNA promoter-driven transcriptional activity when a rRNA promoter is inserted into a non-RNA pol I environment namely outside of the nucleolus. Thus, the target site-specific differences in RNA pol I-driven luciferase activity are in good agreement with the nuclear organization.

The findings in this dissertation demonstrate that GT-rich promoter elements are capable to drive directional transcription initiation. The results here show that the coding strand of TSRs is enriched in GT-rich 10mers and that these are able to promote transcription initiation. The number of identified 10mers is the highest at the 5'-end of TSRs and decreases towards the 3'-end. In concordance, DNA fragments derived from the 5'-end of TSRs yielded the highest luciferase activity. Both findings reflect experimentally the results obtained from the primary transcriptome data and thus further support that transcription is initiated at the 5'-end of TSRs. Moreover, these correlations underline the importance of the identified GT-rich elements for transcription initiation. For several organisms it has been shown that sequences composed of alternating purine-pyrimidine sequences have the potential to adopt a Z-DNA structure (Herbert *et al.*, 1999; Rich and Zhang, 2003). In Z-DNA, the alternating pyrimidines and purines are oriented in *anti*- and *syn*-conformation, respectively, resulting in a zigzag-shaped sugar-phosphate backbone (Figure 6.1; Wang *et al.*, 1979). In B-DNA, they exclusively have an *anti*-conformation. Z-DNA conformation mostly occurs in d(GC)_n repeats, followed by d(TG)_n repeats and d(TA)_n repeats (Wang and Vasquez, 2007). The energetically unfavorable Z-DNA conformation is stabilized by negative supercoiling (Rahmouni and Wells, 1989). In

eukaryotes, the DNA is generally negatively supercoiled, and the energy is absorbed by nucleosome formation to prevent a torsional strain (Ausio *et al.*, 1987). During transcription, RNA polymerase does not rotate on the opened DNA strand resulting in positive supercoiling in front and negative supercoiling behind the progressing polymerase. Thus, active transcription facilitates Z-DNA formation at permissive regions (Liu and Wang, 1987).

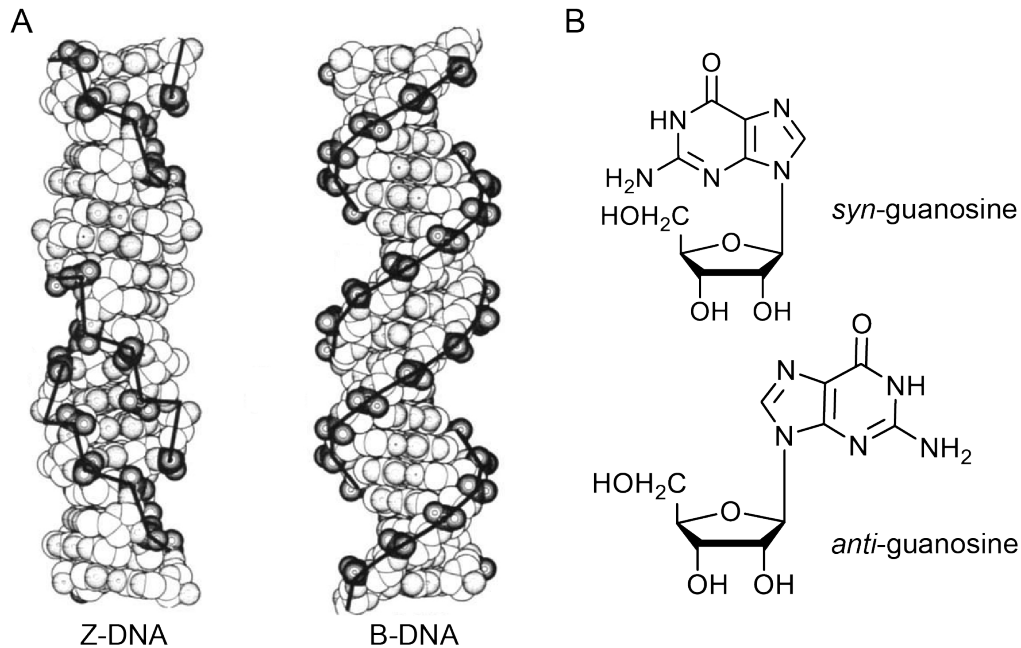


Figure 6.1 The structure of Z-DNA.

(A) van der Waals models of Z-DNA and B-DNA. For Z-DNA, two views are shown. The view on the right represents the double helix shown on the left rotated about 30°. The respective sugar-phosphate backbone is highlighted with a bold line. Modified from (Wang *et al.*, 1979). (B) Structures of *syn*- and *anti*-guanosine nucleoside conformations.

On the single-gene level Z-DNA has been shown to increase the expression level of the investigated genes (Wittig *et al.*, 1992; Liu *et al.*, 2001; Wong *et al.*, 2007; Maruyama *et al.*, 2013). In a genome-wide study in human cells the Z-DNA binding protein Zaa has been demonstrated to bind in promoter regions containing sequences with the highest potential to form Z-DNA. In addition, those regions were found to co-localize with RNA pol II enrichment and histone marks associated with active transcription (Shin *et al.*, 2016). In yeast, Z-DNA has been demonstrated to block nucleosome formation creating an open chromatin state around the TATA-box (Wong *et al.*, 2007). The in this study identified GT-rich promoter elements are composed of GT-rich 10mers found to be enriched on the coding strand of TSRs. Among the most enriched are sequences of alternating Gs and Ts (Appendix Table 7.2). Thus, these sequences can potentially adopt a Z-DNA conformation and the findings in this study demonstrate that they promote transcription initiation and are depleted of nucleosomes. So

far, there is no evidence for the presence of Z-DNA in *T. brucei*. Upon isolation, kinetoplastid DNA has been shown to have the potential to adopt Z-DNA structure (Liu *et al.*, 2006). To investigate whether Z-DNA structures are present in the *T. brucei* genome, prediction tools like Z-hunt and Z-catcher (Schroth *et al.*, 1992; Li *et al.*, 2009b) could be used to find potential Z-DNA forming sequences and IF experiments using polyclonal (Lafer *et al.*, 1981) and monoclonal (Möller *et al.*, 1982) antibodies could be performed.

Data of this study rise evidence that GT-rich elements provide directionality to transcription initiation. The findings demonstrate that the insertion of the reverse complement sequence of the GT-rich promoter element yields a strongly reduced luciferase activity. Additional experiments in which the GT-rich promoter element and its reverse complement sequence, resp., are flanked by *RLUC* and *FLUC* could be performed to further support this hypothesis. In this study *RLUC* has been used to normalize for cell number. Thus, in the setting described above ectopically expressed *lacZ* may be used for this application.

In *T. brucei* it has been suggested that nearby active transcription may influence the transcription level (McAndrew *et al.*, 1998). Since the resistance gene within the targeting construct is transcribed divergently by a T7 promoter located adjacent to the GT-rich promoter element additional experiments were performed to exclude this possibility. To this end, two tetracycline operators (TetO) have been inserted between the T7 promoter and the rRNA promoter or the GT-rich promoter element, resp. (construct pCW24v4 and derivatives). Since all experiments have been performed in the SM background, a cell line that expresses the T7 polymerase and the tetracycline repressor (TetR), putative transcriptional activity from the T7 promoter towards the tested promoter would be blocked by the binding of TetR to TetO. The analysis of these cell lines revealed no influence of the transcriptional activity of T7 on the transcription level driven by the GT-rich promoter element (Figure 4.8).

Taken together, these findings suggest that the link between a lack of well-defined sequence elements and a lack of transcriptional regulation may be highly conserved in evolution.

6.3 GT-rich promoter elements contribute to targeted H2A.Z deposition

It has been long known that the histone variant H2A.Z is enriched within regions of ~10 kb at the 5'-end of PTUs (Siegel *et al.*, 2009). The mechanism contributing to this rather special organization of H2A.Z-containing nucleosomes however, remained elusive to date. Factors homologue to the key players involved in H2A.Z deposition in yeast have been identified in *T. brucei*. Some have been shown to co-localize with H2A.Z, such as the bromodomain-

containing factor BDF3 (Siegel *et al.*, 2009), the PTMs H3K4me3 and H4K10ac (Siegel *et al.*, 2009; Wright *et al.*, 2010). In addition, homologues of histone acetyltransferases (Kawahara *et al.*, 2008; Siegel *et al.*, 2008), several histone methyltransferases (Figueiredo *et al.*, 2009) and several BDFs (Siegel *et al.*, 2009) have been identified. Thus, it is hypothesized, that *T. brucei* utilizes a similar mechanism for H2A.Z deposition. Previous studies performed in yeast demonstrated that an insertion of a 22 bp sequence from the *SNT1* promoter in the center of an inactive gene is sufficient to induce formation of an NDR flanked by two H2A.Z-containing nucleosomes (Raisner *et al.*, 2005).

Indeed, the findings in this study demonstrate that also in *T. brucei* the targeted H2A.Z deposition is DNA sequence-mediated. GT-rich promoter elements capable to initiate transcription were shown to promote the incorporation of H2A.Z to the site of insertion. However, the amount of newly incorporated H2A.Z was lower compared to endogenous H2A.Z levels. These findings could be explained by three different hypotheses: i) The GT-rich element contains a DNA sequence motif that serves as binding site for proteins involved in shaping the chromatin structure, ii) H2A.Z deposition is stimulated by transcription initiation and its associated factors or iii) the GT-rich element establishes a certain chromatin structure promoting H2A.Z deposition. Regarding the first hypothesis, the 22 bp sequence inserted in yeast contained a Reb1 binding site and an adjacent poly(dA:dT)₇ tract. Both have been shown to act redundantly to establish an NDR and H2A.Z deposition (Raisner *et al.*, 2005). The Reb1 binding motif has been shown to be even more conserved among species than the TATA-box (Elemento and Tavazoie, 2005) and is essential for NDR formation in a subset of yeast promoters (Hartley and Madhani, 2009). The here tested sequences are rich in Gs and stretches of Ts, as well. However, so far neither Reb1 homologues have been reported for *T. brucei*, nor it is likely that the synthetic generated GT-rich elements incidentally contain binding motifs. In the experimental setting devised in this work, it is not possible to draw conclusions whether transcription initiation stimulates H2A.Z deposition (second hypothesis). However, findings in yeast provide evidence, that H2A.Z deposition does not require active transcription (Raisner *et al.*, 2005). High-resolution nucleosome mapping of the insertion site of the short GT-rich promoter element indicate that the GT-rich sequence is depleted of nucleosomes (Figure 5.4). A reason for this depletion may be that the promoter element contains sequences that are unfavored by nucleosome formation at a high density, namely intrinsically rigid and repetitive GT-rich sequences and stretches of Ts. Indeed, findings in yeast suggest that NDR establishment is necessary for H2A.Z deposition (Hartley and Madhani, 2009). Thus, the results presented in this work are in line with the third hypothesis.

The low overall nucleosome occupancy in this region may explain why the H2A.Z levels at GT-rich regions did not reach those of endogenous TSRs.

Taken together, in this study specific DNA sequences have been identified that affect local chromatin structure.

6.4 NDRs regulate gene expression post-transcriptionally and are affected by the DNA sequence

In contrast to the conservation of mechanistic details during RNA pol II transcription initiation in *T. brucei* discussed above, the findings in this study revealed significant differences concerning the chromatin structure at transcription initiation sites. While in other eukaryotes promoter regions and splice motifs are depleted of nucleosomes, results from this work demonstrate that only the latter are depleted in *T. brucei*. This suggests that NDRs are associated with mRNA maturation rather than transcription initiation. Given the capability of NDRs to regulate DNA accessibility, the absence of NDRs at TSRs may reflect the lack of transcriptional regulation in *T. brucei*. Thus, they highlight the importance of post-transcriptional mechanisms of gene regulation, such as *trans*-splicing. Genome-wide nucleosome occupancy maps revealed an increase of nucleosome occupancy across exons compared to introns (Schwartz *et al.*, 2009). Splicing occurs co-transcriptionally. Hence, it has been proposed that positioned nucleosomes at the 5'-end of exons may function as 'speed bumps' to slow down RNA pol II elongation and thereby promote the inclusion of exons (Schwartz and Ast, 2010). This model proposes, that exon selection is influenced by the presence or absence of nucleosomes, which thereby affect alternative splicing. The outcome of several studies where SASs have been mapped genome-wide in *T. brucei* suggest that *T. brucei* also utilizes alternative splicing. It was shown that the major SAS of many genes differs between life cycle stages yielding transcripts with different 5'UTRs (Kolev *et al.*, 2010; Nilsson *et al.*, 2010; Siegel *et al.*, 2010). Thus, in the absence of transcriptional regulation, alternative *trans*-splicing has been proposed as a mechanism to regulate gene expression.

The data in this work revealed a strong increase in nucleosome occupancy at the 5'-end of exons. The correlation between nucleosome occupancy and RNA pol II distribution proposes that the increase in nucleosome occupancy may act as a barrier and slows RNA pol II elongation, which results in the observed increase in RNA pol II levels. The decrease in RNA pol II elongation speed may increase co-transcriptional *trans*-splicing efficiency and affect the choice of SASs. Thus, changes in nucleosome positioning between life cycle stages may contribute to the observed life-cycle specific SAS preferences. Within the scope of this

dissertation MNase-ChIP-seq experiments have been performed in the procyclic form of *T. brucei*, but have not been analyzed, yet. Thorough analysis of these data may contribute to confirm this hypothesis.

The DNA sequence affects *trans*-splicing and nucleosome occupancy at polyY tracts. PolyY tracts are of central importance on both, the DNA level and the RNA level. On the one hand, given their homopolymeric nature, these sequences are rigid and thereby unfavored during nucleosome formation resulting in decreased nucleosome occupancy. On the other hand, they serve as binding sites for U2AF65, an important spliceosome subunit. Based on the analysis of nucleosome occupancy in several organisms it has been suggested, that the reason for the observed differences in chromatin structure at exon/intron boundaries lies within the DNA sequence itself. Thus, splicing signals that were previously been thought to act only on the RNA level, may also affect processes on the DNA level, such as the chromatin structure of introns and exons (Schwartz *et al.*, 2009). The in this study observed differences in nucleosome occupancy and gene expression due to distinct compositions of polyY tracts support the above-mentioned hypothesis. However, the polyY tract yielding the strongest depletion induced less gene expression than the endogenous polyY tract. A reason may be a disruption of the U2AF65 binding motif in order to extend the homopolymeric sequence. As expected, no gene expression could be detected in the absence of a polyY tract. However, on the chromatin level, an NDR could still be observed. Hence, additional factors such as chromatin remodelers and sequence motifs within the downstream 5'UTR may be either additionally or redundantly involved in shaping the nucleosomal landscape around exon/intron boundaries.

Taken together, these lines of evidence demonstrate an impact of nucleosome occupancy on RNA maturation in a the highly divergent eukaryotic parasite. The results in this study highlight nucleosome positioning as an additional level of gene expression regulation in *T. brucei* and support the conservation of a link between chromatin structure and regulation of gene expression among eukaryotes.

6.5 Conclusion

Trypanosomatids distinguish themselves from other eukaryotes by their lack of transcriptional regulation, the organization of genes in long PTUs and the apparent absence of promoter motifs. This suggests that trypanosomatids utilize fundamentally distinct mechanisms to regulate gene expression. The findings in this dissertation demonstrate that specific DNA sequence elements can drive directional transcription and affect local chromatin structure.

Furthermore, the data here establish a link between chromatin structure and RNA maturation. Along these lines, these findings suggest that regardless its evolutionary divergence, *T. brucei* utilizes several of the mechanisms found in other eukaryotes to regulate its gene expression. This implies, that many of the strategies are highly conserved in evolution.

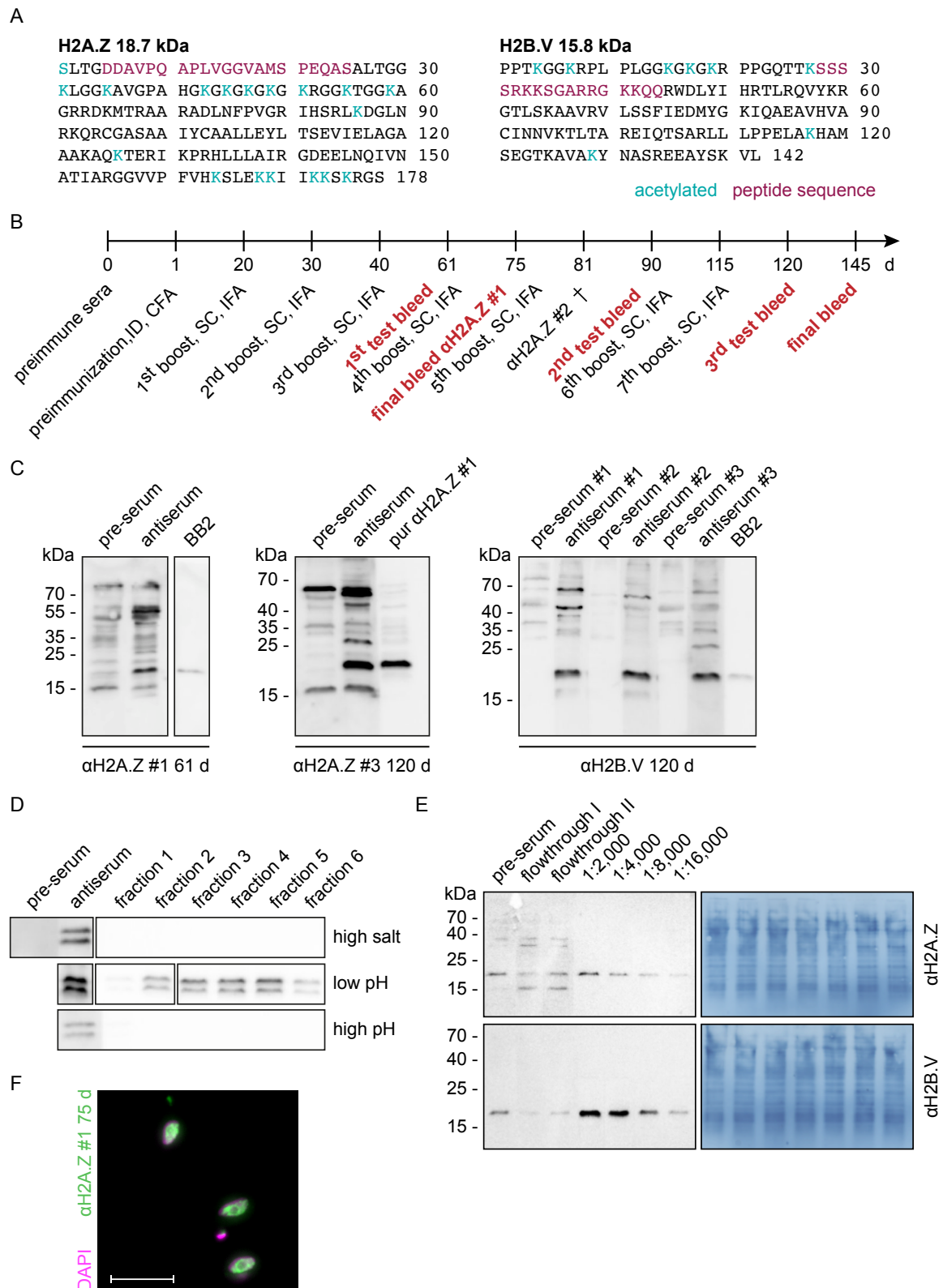
Simultaneously, the findings accentuate the divergence between *T. brucei* and the majority of eukaryotic organisms in terms of nucleosome occupancy at transcription initiation sites. The observed differences may reflect the lack of RNA pol II transcription regulation in *T. brucei* and may support its dependency on post-transcriptional mechanisms of gene expression. Furthermore, these findings open new perspectives on our understanding of mechanistic requirements during transcription initiation in eukaryotes.



Appendix

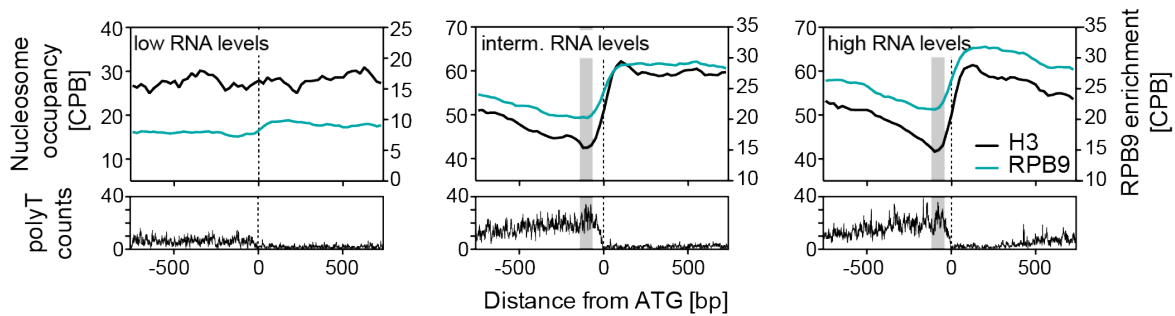
7.1	Appendix Figures	117
7.2	Appendix Tables	119

7.1 Appendix Figures

**Appendix Figure 7.1 Generation of α H2A.Z and α H2B.V.**

(A) Amino acid sequence of H2A.Z (Tb427.07.6360) and H2B.V (Tb427tmp.02.5250). Acetylated lysines were determined previously (Johannes Thürich, unpublished) and are shown in cyan. The sequence of the peptide used for immunization is shown in pink. (B) Timeline of the immunization protocol. ID, intradermal; CFA, Complete Freund's adjuvant; SC, subcutaneous; IFA, Incomplete Freund's adjuvant. (C) Western blot analysis of the first test bleed of α H2A.Z #1 on cell lysates of SM Ty1-H2A.Z H2A.Z^{-/-} (left panel), the third test bleed of α H2A.Z #3 and purified α H2A.Z #1 on cell lysates of Wt (middle panel) and the third test

bleed of α H2B.V #1-3 on cell lysates of BFJEL43 (right panel). Sera were diluted 1:1,000. BB2 (diluted 1:1000) is used as size control for Ty1-H2A.Z and Ty1-H2B.V, respectively. (D) Evaluation of different elution strategies (high salt, low pH, high pH) for α H2A.Z #1 via western blot on cell lysates of SM Ty1-H2A.Z H2A.Z. Elution fractions were diluted 1:250. The antiserum control is shown as loading control. (E) Western blot analysis of purified α H2A.Z #1 (final bleed, upper panels) and α H2B.V #2 (final bleed, lower panels). In total, 10 ml of each antiserum were affinity purified during two rounds of 5 ml each (flowthrough I and II). Several dilutions of the purified antibodies were tested (left panels). The amido black stained nitrocellulose is shown as loading control (right panels). (F) Representative fluorescence microscopy using α H2A.Z #1 75 d 1:500. Scale bar, 7.5 μ m.



Appendix Figure 7.2 Nucleosome depletion correlates with the level of gene expression.

Average nucleosome occupancy (black) and RPB9 enrichment (cyan) plotted relative to the ATG of the 25 % of genes yielding the highest RNA levels (upper left panel, $n = 2,753$), the 25 % of genes yielding intermediate RNA levels (upper middle panel, $n = 2,753$) and the 25 % of genes yielding the lowest RNA levels (upper right panel, $n = 2,753$). RNA levels were determined previously (Vasquez *et al.*, 2014). The number of polyT tracts composed of 10mers of Ts (Hamming distance = 1) within the regions analyzed in the upper panels was counted and plotted relative to the ATG (lower panels). The polyT enrichment is highlighted in grey.

7.2 Appendix Tables

Appendix Table 7.1 List of RNA pol II transcription initiation sites.

Regions were manually chosen based on clear RPB9 and short primary transcript enrichment over background. Coordinates were chosen based on short primary transcript enrichment.

chromosome	start coordinate	end coordinate	strand
Tb427_01_v4	283488	285578	coding (+)
Tb427_01_v4	525915	528005	coding (+)
Tb427_01_v4	654567	656134	noncoding (-)
Tb427_01_v4	655375	656894	coding (+)
Tb427_01_v4	763590	765348	noncoding (-)
Tb427_01_v4	765063	766155	coding (+)
Tb427_01_v4	813539	814917	coding (+)
Tb427_01_v4	1001122	1002309	noncoding (-)
Tb427_02_v4	310446	311556	noncoding (-)
Tb427_02_v4	311052	312262	coding (+)
Tb427_02_v4	513589	514901	noncoding (-)
Tb427_02_v4	515154	516213	coding (+)
Tb427_02_v4	900082	900990	noncoding (-)
Tb427_02_v4	902708	904574	coding (+)
Tb427_02_v4	1027702	1029821	coding (+)
Tb427_03_v4	138226	139444	noncoding (-)
Tb427_03_v4	139039	140404	coding (+)
Tb427_03_v4	203533	204788	noncoding (-)
Tb427_03_v4	204456	206635	coding (+)
Tb427_03_v4	587251	588062	noncoding (-)
Tb427_03_v4	1374695	1376024	coding (+)
Tb427_03_v4	1611691	1612761	coding (+)
Tb427_04_v4	316754	317684	noncoding (-)
Tb427_04_v4	536669	538158	noncoding (-)
Tb427_04_v4	948794	950004	noncoding (-)
Tb427_04_v4	968980	970613	noncoding (-)
Tb427_04_v4	1394537	1396135	noncoding (-)
Tb427_05_v4	318852	321032	coding (+)
Tb427_05_v4	500545	502483	coding (+)
Tb427_05_v4	681753	682721	noncoding (-)
Tb427_05_v4	919594	920966	noncoding (-)
Tb427_05_v4	920563	922743	coding (+)
Tb427_05_v4	1099914	1101771	noncoding (-)
Tb427_05_v4	1100641	1102983	coding (+)
Tb427_05_v4	1333232	1334927	noncoding (-)
Tb427_06_v4	477838	479971	noncoding (-)
Tb427_06_v4	687675	689193	noncoding (-)
Tb427_06_v4	688796	690350	coding (+)
Tb427_06_v4	1279437	1280846	noncoding (-)
Tb427_06_v4	1283921	1285511	coding (+)
Tb427_07_v4	29445	30496	noncoding (-)
Tb427_07_v4	32338	34572	coding (+)
Tb427_07_v4	229887	231026	noncoding (-)
Tb427_07_v4	233876	235453	coding (+)
Tb427_07_v4	349286	351082	coding (+)
Tb427_07_v4	712036	713482	noncoding (-)
Tb427_07_v4	715674	717032	coding (+)
Tb427_07_v4	1014605	1016270	noncoding (-)
Tb427_07_v4	1018550	1020039	coding (+)
Tb427_07_v4	1239900	1241126	noncoding (-)
Tb427_07_v4	1312265	1313185	noncoding (-)
Tb427_07_v4	1314895	1317173	coding (+)
Tb427_07_v4	1748259	1749222	coding (+)
Tb427_07_v4	1905134	1906316	noncoding (-)
Tb427_07_v4	1906361	1907675	coding (+)
Tb427_07_v4	1925428	1926918	noncoding (-)

chromosome	start coordinate	end coordinate	strand
Tb427_07_v4	2143974	2145989	noncoding (-)
Tb427_07_v4	2161506	2162206	coding (+)
Tb427_08_v4	445264	446094	coding (+)
Tb427_08_v4	624090	625197	noncoding (-)
Tb427_08_v4	624894	626195	coding (+)
Tb427_08_v4	876026	877881	noncoding (-)
Tb427_08_v4	876830	878408	coding (+)
Tb427_08_v4	1399543	1400927	noncoding (-)
Tb427_08_v4	1443787	1445919	noncoding (-)
Tb427_08_v4	1449550	1450712	coding (+)
Tb427_08_v4	1609457	1611285	noncoding (-)
Tb427_08_v4	1739942	1741658	coding (+)
Tb427_08_v4	1997891	1998527	coding (+)
Tb427_08_v4	2241988	2243178	coding (+)
Tb427_09_v4	637298	638253	noncoding (-)
Tb427_09_v4	640302	640813	coding (+)
Tb427_09_v4	695435	697106	coding (+)
Tb427_09_v4	831744	833553	noncoding (-)
Tb427_09_v4	833144	834338	coding (+)
Tb427_09_v4	1236301	1237153	noncoding (-)
Tb427_09_v4	1381657	1383431	noncoding (-)
Tb427_09_v4	1516906	1518646	noncoding (-)
Tb427_09_v4	1719955	1721320	noncoding (-)
Tb427_09_v4	1723778	1724972	coding (+)
Tb427_09_v4	2466498	2468272	noncoding (-)
Tb427_09_v4	2467283	2468852	coding (+)
Tb427_10_v5	327642	330012	noncoding (-)
Tb427_10_v5	447781	449630	noncoding (-)
Tb427_10_v5	635099	636301	coding (+)
Tb427_10_v5	1103371	1105221	coding (+)
Tb427_10_v5	1231241	1232860	noncoding (-)
Tb427_10_v5	1628037	1629609	noncoding (-)
Tb427_10_v5	1634514	1635808	coding (+)
Tb427_10_v5	1887018	1888543	noncoding (-)
Tb427_10_v5	1923010	1926340	noncoding (-)
Tb427_10_v5	2059809	2060825	noncoding (-)
Tb427_10_v5	2064111	2065544	coding (+)
Tb427_10_v5	2643554	2644988	noncoding (-)
Tb427_10_v5	2754632	2755973	noncoding (-)
Tb427_10_v5	2758056	2759767	coding (+)
Tb427_10_v5	3168176	3169794	coding (+)
Tb427_10_v5	3312052	3313809	noncoding (-)
Tb427_10_v5	3421417	3422758	noncoding (-)
Tb427_10_v5	3620487	3621781	noncoding (-)
Tb427_10_v5	3623170	3625066	coding (+)
Tb427_10_v5	3946453	3948025	coding (+)
Tb427_11_01_v4	181839	182366	coding (+)
Tb427_11_01_v4	542318	543900	noncoding (-)
Tb427_11_01_v4	545429	546955	coding (+)
Tb427_11_01_v4	890328	891882	noncoding (-)
Tb427_11_01_v4	898826	899936	noncoding (-)
Tb427_11_01_v4	900437	901381	coding (+)
Tb427_11_01_v4	1264918	1266056	noncoding (-)
Tb427_11_01_v4	1267168	1267750	coding (+)
Tb427_11_01_v4	1874030	1875556	noncoding (-)
Tb427_11_01_v4	2001652	2002595	noncoding (-)
Tb427_11_01_v4	2002346	2003123	coding (+)
Tb427_11_01_v4	2188570	2189652	noncoding (-)
Tb427_11_01_v4	2192625	2193291	coding (+)
Tb427_11_01_v4	2598986	2600485	noncoding (-)
Tb427_11_01_v4	2603180	2604401	coding (+)
Tb427_11_01_v4	2994848	2996124	noncoding (-)
Tb427_11_01_v4	3154716	3156714	coding (+)
Tb427_11_01_v4	3450120	3450758	noncoding (-)

chromosome	start coordinate	end coordinate	strand
Tb427_11_01_v4	3594156	3595822	noncoding (-)
Tb427_11_01_v4	3595517	3596627	coding (+)
Tb427_11_01_v4	4025459	4026375	noncoding (-)
Tb427_11_01_v4	4482898	4484897	noncoding (-)
Tb427_11_01_v4	4484287	4485508	coding (+)

Appendix Table 7.2 List of 10mers enriched to at least 6-fold on coding strand compared to the noncoding strand across TSRs.

To identify 10mers enriched on the coding strand compared to the noncoding strand across TSRs, the sequence of each TSR (n = 199) was divided in 5 equally-spaced regions. For each region the number of different 10mers was determined for the coding (blue) and noncoding strand (orange). This list contains only 10mers enriched at least 6-fold on the coding compared to the noncoding strand.

10mer	coding strand region					noncoding strand region				
	1	2	3	4	5	1	2	3	4	5
GGGGGGGGGG	132	7	27	19	18	17	12	3	8	1
GTGTGTGTGG	18	2	4	4	2	1	1	0	1	0
TTTTTGT	18	18	17	6	10	2	2	0	2	3
GGTGTGTGTG	15	0	2	1	2	0	0	0	1	0
CGGGGGGGGG	13	0	3	4	2	1	2	0	1	0
TATTTTTTTT	13	15	9	16	9	2	4	5	8	3
TTTTTAAAT	13	11	7	3	3	2	4	2	4	2
TTTTTTGT	13	11	17	8	12	2	2	2	3	3
TTTTTTTGC	13	21	12	10	8	2	6	5	4	2
TTTTTTTGG	13	6	7	7	2	2	0	2	1	4
GTTTTTGT	13	9	3	0	1	0	1	0	2	1
TGTGTGTGGG	12	0	3	3	0	1	1	0	0	1
TTTTATTTT	12	11	13	11	7	2	4	3	2	1
TTTTGT	12	13	19	13	3	2	3	2	5	2
GTGTGTGGGG	11	0	3	3	0	1	0	1	0	0
TGTGTTGTG	11	6	1	4	2	1	0	1	1	0
TTTGTGTGTG	11	2	2	4	2	1	0	2	1	0
TGTTTTTGT	11	10	3	3	2	0	2	2	2	0
CGTGTGTGTG	10	3	5	4	2	1	0	0	0	2
GGGGGGGGGC	10	2	3	0	4	1	4	0	1	0
TTTGT	10	5	6	3	7	1	2	1	0	0
TTTGT	10	7	4	4	8	1	1	3	4	0
TTCCCTTT	10	3	4	3	7	0	1	2	4	1
GTGTGTGTGC	9	3	2	1	3	1	0	1	1	2
GTGTGTTGT	9	3	1	2	1	1	1	1	1	0
TATGTGTGTG	9	5	4	1	1	1	0	3	0	0
TGCGTGCGTG	9	0	2	2	1	1	1	0	0	1
TTATTTTT	9	14	12	11	11	1	2	3	3	1
TTTGT	9	3	3	3	5	1	1	2	1	1
TTTTTTTT	9	9	10	7	4	1	0	4	6	2
GTGCGTGCGT	9	0	2	1	1	0	2	1	0	1
GTTGTTGT	9	2	1	1	1	0	2	0	1	0
TGTATGTGTG	9	4	3	3	2	0	2	2	1	0
TGTGTGTGTT	9	1	2	2	3	0	0	1	1	0
ATGTTTTTT	8	4	7	4	5	1	2	2	0	0
CCCCCTCCC	8	3	3	1	1	1	1	1	4	6
CCTTTCTTT	8	4	5	3	0	1	5	3	2	2
GCGTGTGTGT	8	3	4	1	3	1	1	0	0	1
GGGGGAAAC	8	0	1	0	1	0	1	2	0	0
GGTTGTTGTT	8	5	1	3	3	0	1	0	0	0
GTTTGTGTGT	8	2	3	1	0	0	0	1	0	0
TCATTTTTT	8	10	8	2	4	0	2	5	1	0
TCGTTTTTT	8	5	8	6	0	0	2	0	1	0
TGCGTGTGTG	8	2	7	2	3	0	1	0	0	1
TGGTGTGTGT	8	2	2	2	0	0	0	2	1	0
TGTTTGTGTG	8	5	3	4	3	0	1	1	1	0
TTGCTTTTT	8	5	4	4	2	0	1	1	4	2
TTTTGTTC	8	7	2	4	6	0	1	1	3	1

10mer	coding strand region					noncoding strand region				
	1	2	3	4	5	1	2	3	4	5
AATGTTTTTT	7	5	1	3	1	1	1	2	1	0
ATGTGTGTGT	7	4	4	2	4	1	1	2	2	1
ATTTATTTT	7	5	3	4	5	1	2	4	3	4
CCCTTTTCTT	7	2	2	4	3	1	1	1	2	0
CGAAGAAAAT	7	1	2	2	1	1	1	1	0	2
CGGTGGTGT	7	0	2	1	0	1	0	0	1	1
GGGGGGGGAG	7	1	4	4	4	1	1	1	0	1
GTCTTTTTTT	7	2	1	1	3	1	2	0	1	3
GTGGGTGCGG	7	0	1	0	2	1	0	0	0	0
TCTTTCTCCC	7	0	1	0	0	1	2	2	1	0
TGTTGCTGCT	7	1	2	3	0	1	1	0	1	0
TGTTTTTGT	7	3	7	11	3	1	0	2	2	3
TGTTTTTCT	7	2	5	3	7	1	2	0	1	0
TTGTTTTTC	7	4	2	5	5	1	1	0	1	1
TTTCCCCTT	7	4	4	2	6	1	2	3	6	0
TTTGCATTT	7	1	2	2	1	1	1	1	3	1
TTTGTGAAA	7	0	2	0	0	1	0	0	2	0
TTTTTTGGG	7	1	5	2	1	1	0	0	0	1
AGGGGGAGAG	7	0	0	0	0	0	0	1	0	1
GGTGGGTGGT	7	0	0	0	0	0	0	0	1	0
GTGCGTGTGT	7	2	3	3	1	0	3	0	0	0
GTGTGTGGAG	7	0	1	2	0	0	0	1	1	0
GTGTTGTGG	7	0	1	1	1	0	0	0	0	0
GTTTTTTTGT	7	4	5	3	1	0	1	3	0	0
TCTTTGTTTT	7	5	4	3	7	0	1	0	1	0
TGCTTTTTTC	7	1	1	2	1	0	0	0	1	0
TGTGTATGTG	7	3	4	4	3	0	1	1	1	0
TGTGTGTTG	7	4	0	2	1	0	0	4	0	0
TGTTGTTTT	7	3	2	1	0	0	0	1	0	1
TTATTTTTGT	7	8	3	4	3	0	1	2	0	2
TTCTTTTTG	7	2	2	1	2	0	2	0	0	3
TTCTTTTTT	7	5	6	7	0	0	2	0	1	0
TTGTTTTGTT	7	6	6	1	5	0	5	0	1	0
TTTCGTTTT	7	5	3	4	0	0	1	2	1	1
AAAAAAGGAG	6	1	2	2	0	1	2	1	0	1
AAAAAGAAAG	6	2	5	3	2	1	5	7	7	6
ACCGCTGCTG	6	2	1	2	0	1	0	1	0	1
CGGTGGAGGC	6	0	0	0	1	1	0	0	0	0
GAAGAAAATT	6	3	1	1	0	1	1	1	2	2
GCTGCTGCTT	6	1	2	0	3	1	1	0	0	0
GGTGCTGCTG	6	0	2	2	2	1	2	0	0	0
GTATATATTT	6	5	4	5	0	1	1	0	3	2
GTGGGTGGTG	6	0	2	0	1	1	0	0	1	0
TCCCCTTTC	6	1	0	2	2	1	0	0	3	1
TGGGCACGTG	6	0	0	0	0	1	0	1	1	0
TGGGTGGTGG	6	2	2	1	1	1	0	0	0	0
TGTGTGGGGG	6	0	1	2	1	1	0	0	0	0
TGTGTTTTTT	6	4	6	2	0	1	0	0	0	1
TTCATTTTTT	6	8	3	3	4	1	3	5	2	1
TTGTGTGTGT	6	1	1	4	1	1	0	1	1	0
TTGTGTGTTT	6	4	1	1	1	1	0	2	2	0
TTGTTTTTTT	6	4	9	6	6	1	4	3	5	1
TTTCTTCTT	6	6	5	7	3	1	2	1	3	2
TTTTCTATTT	6	4	2	2	2	1	3	0	0	0
TTTTCTGAGG	6	0	0	1	0	1	1	0	0	0
TTTTTTGTCA	6	4	6	2	0	1	1	0	0	0
TTTTTTTCTA	6	4	6	2	6	1	2	3	0	1
TTTTTTTTTN	6	3	0	3	3	1	2	1	1	0
CAGTGGAAAA	6	1	0	3	2	0	2	0	2	1
CATTGTTTGT	6	1	2	0	1	0	0	1	0	0
GAAGCGGTGG	6	0	1	0	0	0	0	0	0	0
GAGGGGGAGA	6	1	0	1	0	0	0	1	0	0
GGAAACTCTC	6	0	0	0	0	0	0	1	1	0
GGCTGATGGC	6	0	1	0	1	0	0	0	0	0
GGGCGTGGGG	6	0	0	0	0	0	0	0	0	1
GGGGGGGGCT	6	2	1	0	2	0	0	0	2	0
GGGTGTTGTG	6	0	0	1	1	0	0	1	0	0

10mer	coding strand region					noncoding strand region				
	1	2	3	4	5	1	2	3	4	5
GGTGATGGCT	6	0	0	0	1	0	0	0	0	0
GGTGTGTTT	6	1	3	0	2	0	0	1	0	1
GTATGTGTGT	6	4	4	4	1	0	2	3	1	0
GTGATGGCTG	6	0	0	1	0	0	0	0	0	0
GTGTGTGGCG	6	0	0	0	1	0	1	0	0	1
GTGTGTGGTG	6	0	2	3	1	0	1	0	0	0
GTGTTGTTTT	6	2	6	1	1	0	3	2	1	1
GTGTTTGTGT	6	3	1	4	2	0	0	2	0	0
GTTGTTGTGG	6	0	0	1	1	0	0	0	0	0
GTTTTATTTT	6	5	2	4	1	0	1	1	2	2
TAGCATCTCA	6	0	0	1	1	0	0	0	0	0
TCGGCTGCGG	6	0	0	1	0	0	1	0	0	0
TGATTGTTTG	6	0	2	0	0	0	0	0	0	0
TGAAAAGGA	6	0	1	1	2	0	1	1	0	0
TGGGGCGTGG	6	0	0	0	0	0	0	0	0	0
TGGTGGTTGC	6	1	0	1	0	0	0	0	0	0
TGTGTGTATG	6	2	2	1	0	0	1	1	1	1
TGTGTGTGAA	6	2	1	1	1	0	0	0	1	0
TGTGTGTGGA	6	2	3	2	0	0	0	0	0	0
TGTTATTTGT	6	2	1	0	1	0	1	0	2	0
TGTTGTGGTT	6	0	1	1	1	0	0	1	0	0
TGTTTGCTTT	6	3	2	1	1	0	1	0	1	2
TGTTTTGTCT	6	1	3	0	1	0	0	0	0	0
TGTTTTGTTT	6	5	6	3	5	0	3	0	2	1
TTATTTTATT	6	6	4	3	1	0	0	1	0	1
TTCGCCCTTT	6	1	0	1	0	0	1	0	0	0
TTGTATGTGT	6	2	1	0	1	0	0	1	0	0
TTGTTGTGGT	6	0	0	1	2	0	0	1	0	0
TTGTTGTTTG	6	2	2	1	1	0	0	2	1	0
TTTATTTTGT	6	3	7	6	2	0	1	1	1	2
TTTCGGTGCT	6	1	1	1	0	0	2	0	0	1
TTTCTCCCTT	6	3	4	2	4	0	2	1	0	2
TTTGCTTTTT	6	4	6	6	2	0	0	2	2	2
TTTGGGGGGG	6	1	6	3	1	0	1	1	1	1
TTTGTATGTG	6	1	1	0	1	0	0	1	1	1
TTTGTGTGTT	6	6	3	1	0	0	0	0	2	0
TTTGTTCCT	6	2	3	3	3	0	1	0	3	2
TTTGTTTTCC	6	2	6	2	4	0	1	1	1	3
TTTTGCCACT	6	0	2	2	0	0	1	0	0	1
TTTTGGGGGG	6	2	5	4	2	0	0	0	1	1
TTTTTTTATT	6	7	9	4	7	0	0	6	1	1
TTTTTTTGAT	6	3	8	3	1	0	0	1	1	0
TTTTTTTGCA	6	6	6	3	1	0	1	2	1	1

Appendix Table 7.3 Information about sequencing data discussed in this study.

The data have been deposited in NCBI's Gene Expression Omnibus (Edgar *et al.*, 2002) and are accessible through GEO Series accession number GSE98061 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98061>). Information about their processing is listed in Appendix Table 7.4.

description	sequencing						
	fastq name	internal library name	type of seq	sequencer	sequencing mode	readlength [bp]	sequenced reads
H2A.Z ChIP	Wt_H2AZ_ChIP_R1.fq.bz2; Wt_H2AZ_ChIP_R2.fq.bz2	NS092	ChIP-Seq	Illumina HiSeq 2500	paired-end	2x100	19307030
TY-RPB9 ChIP	TY-RPB9_input_R1.fq.gz; TY-RPB9_input_R2.fq.gz; TY-RPB9_ChIP_R1.fq.gz; TY-RPB9_ChIP_R2.fq.gz	NS299; NS300	ChIP-Seq	Illumina NextSeq	paired-end	2x76	2000698; 2389665
small 3P-RNA	3P-RNA_PP_R1.fq.bz2; 3P-RNA_PP_R1.fq.bz2	NS279; NS280	RNA-Seq	Illumina NextSeq	paired-end	2x76	19023800; 15803268
GT_210_nt_8dpt	GT_210_nt_input_8dpt_R1.fq.bz2; GT_210_nt_input_8dpt_R2.fq.bz2; GT_210_nt_ChIP_8dpt_R1.fq.bz2; GT_210_nt_ChIP_8dpt_R2.fq.bz2	NS341; NS342	ChIP-Seq	Illumina NextSeq	paired-end	2x76	4389183; 4822644
GT_210_nt_21dpt	GT_210_nt_input_21dpt_R1.fq.bz2; GT_210_nt_input_21dpt_R2.fq.bz2; GT_210_nt_ChIP_21dpt_R1.fq.bz2; GT_210_nt_ChIP_21dpt_R2.fq.bz2	NS230; NS231	ChIP-Seq	Illumina NextSeq	paired-end	2x76	3590078; 3766157
GT_416_nt_8dpt	GT_416_nt_input_8dpt_R1.fq.bz2; GT_416_nt_input_8dpt_R2.fq.bz2; GT_416_nt_ChIP_8dpt_R1.fq.bz2; GT_416_nt_ChIP_8dpt_R2.fq.bz2	NS345; NS346	ChIP-Seq	Illumina NextSeq	paired-end	2x76	5468941; 5380736
GT_416_nt_97dpt	GT_416_nt_input_97dpt_R1.fq.bz2; GT_416_nt_input_97dpt_R2.fq.bz2; GT_416_nt_ChIP_97dpt_R1.fq.bz2; GT_416_nt_ChIP_97dpt_R2.fq.bz2	NS248; NS249	ChIP-Seq	Illumina NextSeq	paired-end	2x76	11750987; 2475432
H3 nucleosome positioning	H3_ChIP-seq_R1.fq.gz; H3_ChIP-seq_R2.fq.gz	NS031	ChIP-Seq	Illumina HiSeq 2500	paired-end	2x100	12615738
H3 ChIP endog. polyY	H3_ChIP_GT_210_nt_R1.fq.gz; H3_ChIP_GT_210_nt_R2.fq.gz	NS273	ChIP-Seq	Illumina NextSeq	paired-end	2x76	23187157
H3 ChIP strong polyY	H3_ChIP_strongPolyY_R1.fq.gz; H3_ChIP_strongPolyY_R2.fq.gz	NS274	ChIP-Seq	Illumina NextSeq	paired-end	2x76	27659287
H3 ChIP no polyY	H3_ChIP_noPolyY_R1.fq.gz; H3_ChIP_noPolyY_R2.fq.gz	NS275	ChIP-Seq	Illumina NextSeq	paired-end	2x76	27946008
H3.V	Siegel et al., 2009	N/A	ChIP-Seq	Siegel et al., 2009	single-end	36	6812684
mRNA Seq	Vasquez et al., 2014	NS025	RNA-Seq	Illumina HiSeq 2500	single-end	2x100	20213835

Appendix Table 7.4 Information about the processing of the sequencing data discussed in this study.

The computational data analysis was implemented as a Unix shell script, which together with further programs generated for this study are available at <https://doi.org/10.5281/zenodo.438156> (DOI:10.5281/zenodo.438156). The processing of aligned reads was performed using a custom-made pipeline, COVERnant version 0.3.0, which is available for download at GitHub (<https://github.com/konrad/COVERnant>). To investigate nucleosome positioning reads were aligned using bowtie version 1.1.1 and processing was performed using the nucwave pipeline (Quintales *et al.*, 2015).

figure	description	mapping of sequenced reads				processing of aligned reads		
		mapping software	mapping parameters	read alignments	reference genome	wig file ws ss	average file ws ss	statistics
3.1C	H2A.Z ChIP	bowtie2 version 2.1.0	local	93,28%	Tb427v24	ws 101 ss 101	-	-
3.1C	H2A.Z ChIP	bowtie2 version 2.1.0	local	93,28%	Tb427v24	ws 1 ss 1	ws 101 ss 101	median
3.1C	TY-RPB9 ChIP	bowtie2 version 2.1.0	local	89,73%	Tb427v24	ws 101 ss 25	-	-
3.2B	H2A.Z ChIP	bowtie2 version 2.1.0	local	93,28%	Tb427v24	ws 101 ss 101	-	-
3.2B	H2A.Z ChIP	bowtie2 version 2.1.0	local	93,28%	Tb427v24	ws 1 ss 1	ws 101 ss 101	median
3.2B	small 3P-RNA	bowtie2 version 2.1.0	R1.fq only, local	46.51%; 51.76%	Tb427v24	ws 1001 ss 25, ratio, forward/reverse	-	-
3.2B	small 3P-RNA	bowtie2 version 2.1.0	R1.fq only, local	46.51%; 51.76%	Tb427v24	ws 1001 ss 25, ratio, forward/reverse	ws 101 ss 101	median
3.5A	H2A.Z ChIP	bowtie2 version 2.1.0	local	93,28%	Tb427v24	ws 101 ss 101	-	-
3.5A	H3 nucleosome positioning	bowtie2 version 2.1.0	local, no-mixed, no-discordant, unique, 137-157	25,18%	Tb427v24	ws 11 ss 11	-	-
3.5A	H3 nucleosome positioning	bowtie2 version 2.1.0	local, no-mixed, no-discordant, unique, 100-130	26,78%	Tb427v24	ws 11 ss 11	-	-
3.5A	H3 nucleosome positioning	bowtie2 version 2.1.0	local, no-mixed, no-discordant, unique, >175	16,29%	Tb427v24	ws 11 ss 11	-	-
3.5B	H2A.Z ChIP	bowtie2 version 2.1.0	local	93,28%	Tb427v24	ws 1 ss 1	ws 101 ss 101	mean
3.5B	H3 nucleosome positioning	bowtie2 version 2.1.0	local, no-mixed, no-discordant, unique, 137-157	25,18%	Tb427v24	ws 1 ss 1	ws 11 ss 11	mean
3.5B	H3 nucleosome positioning	bowtie2 version 2.1.0	local, no-mixed, no-discordant, unique, 100-130	26,78%	Tb427v24	ws 1 ss 1	ws 11 ss 11	mean
3.5B	H3 nucleosome positioning	bowtie2 version 2.1.0	local, no-mixed, no-discordant, unique, >175	16,29%	Tb427v24	ws 1 ss 1	ws 11 ss 11	mean
4.1	H2A.Z ChIP	bowtie2 version 2.1.0	local	93,28%	Tb427v24	ws 101 ss 101	-	-
4.1	small 3P-RNA	bowtie2 version 2.1.0	R1.fq only, local	46.51%; 51.76%	Tb427v24	ws 1001 ss 25, ratio, forward/reverse	-	-
4.2	H2A.Z ChIP	bowtie2 version 2.1.0	local	93,28%	Tb427v24	ws 101 ss 101	-	-
4.2	H3.V	bowtie2 version 2.1.0	local	38,74%	Tb427v24	ws 501 ss 501	-	-
4.2	mRNA Seq	bowtie2 version 2.1.0	local	47,14%	Tb427v24	ws 101 ss 101	-	-
4.3A/B/C	H2A.Z ChIP	bowtie2 version 2.1.0	local	93,28%	Tb427v24	ws 101 ss 101	-	-
4.3A/B/C	H3.V	bowtie2 version 2.1.0	local	38,74%	Tb427v24	ws 501 ss 501	-	-
4.3A/B/C	mRNA Seq	bowtie2 version 2.1.0	local	47,14%	Tb427v24	ws 101 ss 101	-	-
4.4	H2A.Z ChIP	bowtie2 version 2.1.0	local	93,28%	Tb427v24	ws 101 ss 101	-	-
4.4	small 3P-RNA	bowtie2 version 2.1.0	R1.fq only, local	46.51%; 51.76%	Tb427v24	ws 1001 ss 25, ratio, forward/reverse	-	-
4.5A	H2A.Z ChIP	bowtie2 version 2.1.0	local	93,28%	Tb427v24	ws 1 ss 1	ws 101 ss 101	median
4.7	H2A.Z ChIP	bowtie2 version 2.1.0	local	93,28%	Tb427v24	ws 101 ss 101	-	-
4.8A	GT_210_nt_8dpt	bowtie2 version 2.1.0	local	99.62%; 99.21%	Tb427v24_GT_210_nt_8dpt	ws 1 ss 1, ratio	-	-
4.8A	GT_210_nt_21dpt	bowtie2 version 2.1.0	local	82.34%; 91.82%	Tb427v24_GT_210_nt_21dpt	ws 1 ss 1, ratio	-	-
4.8A	GT_416_nt_8dpt	bowtie2 version 2.1.0	local	99.65%; 99.59%	Tb427v24_GT_416_nt_8dpt	ws 1 ss 1, ratio	-	-
4.8A	GT_416_nt_97dpt	bowtie2 version 2.1.0	local	75.19%; 94.17%	Tb427v24_GT_416_nt_97dpt	ws 1 ss 1, ratio	-	-
5.1B	H3 nucleosome positioning	bowtie2 version 2.1.0	local, no-mixed, no-discordant, unique, 137-157	25,18%	Tb927v24	ws 1 ss 1	ws 101 ss 25	median
5.2	H3 nucleosome positioning	bowtie2 version 2.1.0	local, no-mixed, no-discordant, unique, 137-157	25,18%	Tb927v24	ws 1 ss 1	ws 101 ss 25	median
5.2	TY-RPB9 ChIP	bowtie2 version 2.1.0	local	78.95%; 89.73%	Tb927v24	ws 1 ss 1, ratio	ws 101 ss 25	median
5.4	H3 ChIP endog. polyY	bowtie version 1.1.1	--suppress 1,6,7,8 --fr -m 1 -v 2	37,37%	Tb427v24	nucwave default	-	-
5.4	H3 ChIP strong polyY	bowtie version 1.1.1	--suppress 1,6,7,8 --fr -m 1 -v 2	33,07%	Tb427v24	nucwave default	-	-
5.4	H3 ChIP no polyY	bowtie version 1.1.1	--suppress 1,6,7,8 --fr -m 1 -v 2	35,82%	Tb427v24	nucwave default	-	-
7.2	H3 nucleosome positioning	bowtie2 version 2.1.0	local, no-mixed, no-discordant, unique, 137-157	25,18%	Tb927v24	ws 1 ss 1	ws 101 ss 25	median
7.2	TY-RPB9 ChIP	bowtie2 version 2.1.0	local	78.95%; 89.73%	Tb927v24	ws 1 ss 1, ratio	ws 101 ss 25	median

References

- Abbott DW, Ivanova VS, Wang X, Bonner WM, Ausio J (2001) Characterization of the stability and folding of H2A.Z chromatin particles: implications for transcriptional activation. *J Biol Chem*, **276**: 41945–41949
- Adelman K, Lis JT (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet*, **13**: 720–731
- Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, Pugh BF (2007) Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature*, **446**: 572–576
- Allan J, Hartman PG, Crane-Robinson C, Aviles FX (1980) The structure of histone H1 and its location in chromatin. *Nature*, **288**: 675–679
- Alsford S, Horn D (2004) Trypanosomatid histones. *Mol Microbiol*, **53**: 365–372
- Alsford S, Horn D (2011) Elongator protein 3b negatively regulates ribosomal DNA transcription in african trypanosomes. *Mol Cell Biol*, **31**: 1822–1832
- Alsford S, Kawahara T, Glover L, Horn D (2005) Tagging a *T. brucei* RRNA locus improves stable transfection efficiency and circumvents inducible expression position effects. *Mol Biochem Parasitol*, **144**: 142–148
- Aurrecochea C, Barreto A, Brestelli J, Brunk BP, Cade S, Doherty R, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Hu S, Iodice J, Kissinger JC, Kraemer ET, Li W, Pinney DF, Pitts B, Roos DS, Srinivasamoorthy G, Stoeckert CJ, Wang H, Warrenfeltz S (2013) EuPathDB: the eukaryotic pathogen database. *Nucleic Acids Res*, **41**: D684–91
- Ausio J, Zhou G, van Holde K (1987) A reexamination of the reported B→Z DNA transition in nucleosomes reconstituted with poly(dG-m5dC).poly(dG-m5dC). *Biochemistry*, **26**: 5595–5599
- Bangs JD, Crain PF, Hashizume T, McCloskey JA, Boothroyd JC (1992) Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *J Biol Chem*, **267**: 9805–9815
- Bao Y, Shen X (2007) INO80 subfamily of chromatin remodeling complexes. *Mutat Res*, **618**: 18–29
- Basehoar AD, Zanton SJ, Pugh BF (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell*, **116**: 699–709
- Bastin P, Bagherzadeh Z, Matthews KR, Gull K (1996) A novel epitope tag system to study protein targeting and organelle biogenesis in *Trypanosoma brucei*. *Mol Biochem Parasitol*, **77**: 235–239
- Batsché E, Yaniv M, Muchardt C (2006) The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol*, **13**: 22–29
- Bayele HK (2009) *Trypanosoma brucei*: a putative RNA polymerase II promoter. *Exp Parasitol*, **123**: 313–318
- Belotserkovskaya R, Oh S, Bondarenko VA, Orphanides G, Studitsky VM, Reinberg D (2003) FACT facilitates transcription-dependent nucleosome alteration. *Science*, **301**: 1090–1093
- Ben Amar MF, Jefferies D, Pays A, Bakalara N, Kendall G, Pays E (1991) The actin gene promoter of *Trypanosoma brucei*. *Nucleic Acids Res*, **19**: 5857–5862
- Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UC, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DM, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabinowitsch E, Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CM, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El-Sayed NM (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science*, **309**: 416–422
- Biebinger S, Rettenmaier S, Flaspohler J, Hartmann C, Peña-Díaz J, Wirtz LE, Hotz HR, Barry JD, Clayton C (1996) The PARP promoter of *Trypanosoma brucei* is developmentally regulated in a chromosomal context. *Nucleic Acids Res*, **24**: 1202–1211
- Biswas D, Dutta-Biswas R, Stillman DJ (2007) Chd1 and yFACT act in opposition in regulating transcription. *Mol Cell Biol*, **27**: 6279–6287
- Blumenthal T, Evans D, Link CD, Guffanti A, Lawson D, Thierry-Mieg J, Thierry-Mieg D, Chiu WL, Duke K, Kiraly M, Kim SK (2002) A global analysis of *Caenorhabditis elegans* operons. *Nature*, **417**: 851–854
- Brenndorfer M, Boshart M (2010) Selection of reference genes for mRNA quantification in *Trypanosoma brucei*. *Mol Biochem Parasitol*, **172**: 52–55
- Briggs SD, Xiao T, Sun ZW, Caldwell JA, Shabanowitz J, Hunt DF, Allis CD, Strahl BD (2002) Gene silencing: trans-histone regulatory pathway in chromatin. *Nature*, **418**: 498
- Brogaard K, Xi L, Wang JP, Widom J (2012) A map of nucleosome positions in yeast at base-pair resolution. *Nature*, **486**: 496–501
- Bruce A, Alexander J, Julian L, David M, Martin R, Keith R, Peter W (2014) *Molecular Biology of the Cell*, Sixth Edition.
- Burge CB, Tuschl T, MONOGRAPH ... PAS-COLDSRINGHARBOR, 1999 Splicing of

- precursors to mRNAs by the spliceosomes. *Citeseer*.
- Cairns BR (2007) Chromatin remodeling: insights and intrigue from single-molecule studies. *Nat Struct Mol Biol*, **14**: 989–996
- Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engström PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond SM, Wells CA, Orlando V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA, Hayashizaki Y (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet*, **38**: 626–635
- Chang GS, Noegel AA, Mavrich TN, Muller R, Tomsho L, Ward E, Felder M, Jiang C, Eichinger L, Glockner G, Schuster SC, Pugh BF (2012) Unusual combinatorial involvement of poly-A/T tracts in organizing genes and chromatin in *Dictyostelium*. *Genome Res*, **22**: 1098–1106
- Chen W, Luo L, Zhang L (2010) The organization of nucleosomes around splice sites. *Nucleic Acids Res*, **38**: 2788–2798
- Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, Casero D, Bernal M, Huijser P, Clark AT, Krämer U, Merchant SS, Zhang X, Jacobsen SE, Pellegrini M (2010) Relationship between nucleosome positioning and DNA methylation. *Nature*, **466**: 388–392
- Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem*, **78**: 273–304
- Clayton CE (2002) Life without transcriptional control? From fly to man and back again. *EMBO J*, **21**: 1881–1888
- Cole HA, Howard BH, Clark DJ (2012) Genome-wide mapping of nucleosomes in yeast using paired-end sequencing. *Methods Enzymol*, **513**: 145–168
- Cooper SJ, Trinklein ND, Anton ED, Nguyen L, Myers RM (2006) Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res*, **16**: 1–10
- Cordingley JS (1985) Nucleotide sequence of the 5S ribosomal RNA gene repeat of *Trypanosoma brucei*. *Mol Biochem Parasitol*, **17**: 321–330
- Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, **322**: 1845–1848
- Crick F (1970) Central dogma of molecular biology. *Nature*, **227**: 561–563
- Cross GA (1975) Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. *Parasitology*, **71**: 393–417
- Cross GA, Kim HS, Wickstead B (2014) Capturing the variant surface glycoprotein repertoire (the VSGnome) of *Trypanosoma brucei* Lister 427. *Mol Biochem Parasitol*, **195**: 59–73
- Damelin M, Simon I, Moy TI, Wilson B, Komili S, Tempst P, Roth FP, Young RA, Cairns BR, Silver PA (2002) The genome-wide localization of Rsc9, a component of the RSC chromatin-remodeling complex, changes in response to stress. *Mol Cell*, **9**: 563–573
- Das A, Bellofatto V (2003) RNA polymerase II-dependent transcription in trypanosomes is associated with a SNAP complex-like transcription factor. *Proc Natl Acad Sci U S A*, **100**: 80–85
- Das A, Bellofatto V (2009) The non-canonical CTD of RNAP-II is essential for productive RNA synthesis in *Trypanosoma brucei*. *PLoS One*, **4**: e6959
- Das A, Zhang Q, Palenchar JB, Chatterjee B, Cross GA, Bellofatto V (2005) Trypanosomal TBP functions with the multisubunit transcription factor tSNAP to direct spliced-leader RNA gene expression. *Mol Cell Biol*, **25**: 7314–7322
- de la Serna IL, Ohkawa Y, Imbalzano AN (2006) Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. *Nat Rev Genet*, **7**: 461–473
- De Lange T, Borst P (1982) Genomic environment of the expression-linked extra copies of genes for surface antigens of *Trypanosoma brucei* resembles the end of a chromosome. *Nature*, **299**: 451–453
- Dean S, Sunter J, Wheeler RJ, Hodgkinson I, Gluenz E, Gull K (2015) A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids. *Open Biol*, **5**: 140197
- Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. *Genes Dev*, **25**: 1010–1022
- Dechering KJ, Cuelenaere K, Konings RN, Leunissen JA (1998) Distinct frequency-distributions of homopolymeric DNA tracts in different genomes. *Nucleic Acids Res*, **26**: 4056–4062
- Deng W, Roberts SG (2005) A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev*, **19**: 2418–2423
- Dover J, Schneider J, Tawiah-Boateng MA, Wood A, Dean K, Johnston M, Shilatifard A (2002) Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. *J Biol Chem*, **277**: 28368–28371
- Drew HR, Travers AA (1985) DNA bending and its relation to nucleosome positioning. *J Mol Biol*, **186**: 773–790
- Durant M, Pugh BF (2007) NuA4-directed chromatin transactions throughout the *Saccharomyces cerevisiae* genome. *Mol Cell Biol*, **27**: 5327–5335
- Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*, **30**: 207–210
- Elemento O, Tavazoie S (2005) Fast and systematic genome-wide discovery of conserved regulatory elements using a non-alignment based approach. *Genome Biol*, **6**: R18

- Fan X, Lamarre-Vincent N, Wang Q, Struhl K (2008) Extensive chromatin fragmentation improves enrichment of protein binding sites in chromatin immunoprecipitation experiments. *Nucleic Acids Res*, **36**: e125
- Fazio TG, Kooperberg C, Goldmark JP, Neal C, Basom R, Delrow J, Tsukiyama T (2001) Widespread collaboration of Isw2 and Sin3-Rpd3 chromatin remodeling complexes in transcriptional repression. *Mol Cell Biol*, **21**: 6450–6460
- Fernandes AP, Nelson K, Beverley SM (1993) Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: perspectives on the age and origins of parasitism. *Proc Natl Acad Sci U S A*, **90**: 11608–11612
- Figueiredo LM, Cross GA, Janzen CJ (2009) Epigenetic regulation in African trypanosomes: a new kid on the block. *Nat Rev Microbiol*, **7**: 504–513
- Flanagan JF, Mi LZ, Chruszcz M, Cymborowski M, Clines KL, Kim Y, Minor W, Rastinejad F, Khorasanizadeh S (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature*, **438**: 1181–1185
- Ford E, Nikopoulou C, Kokkalis A, Thanos D (2014) A method for generating highly multiplexed ChIP-seq libraries. *BMC Res Notes*, **7**: 312
- Freistadt MS, Cross GA, Branch AD, Robertson HD (1987) Direct analysis of the mini-exon donor RNA of *Trypanosoma brucei*: detection of a novel cap structure also present in messenger RNA. *Nucleic Acids Res*, **15**: 9861–9879
- Freistadt MS, Cross GA, Robertson HD (1988) Discontinuously synthesized mRNA from *Trypanosoma brucei* contains the highly methylated 5' cap structure, m⁷GpppA^{*}A^{*}C(2'-O)mU^{*}A. *J Biol Chem*, **263**: 15071–15075
- Frith MC, Valen E, Krogh A, Hayashizaki Y, Carninci P, Sandelin A (2008) A code for transcription initiation in mammalian genomes. *Genome Res*, **18**: 1–12
- Fussner E, Strauss M, Djuric U, Li R, Ahmed K, Hart M, Ellis J, Bazett-Jones DP (2012) Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. *EMBO Rep*, **13**: 992–996
- Gan L, Ladinsky MS, Jensen GJ (2013) Chromatin in a marine picoeukaryote is a disordered assemblage of nucleosomes. *Chromosoma*, **122**: 377–386
- Ganguli D, Chereji RV, Iben JR, Cole HA, Clark DJ (2014) RSC-dependent constructive and destructive interference between opposing arrays of phased nucleosomes in yeast. *Genome Res*, **24**: 1637–1649
- Gassen A, Brechtefeld D, Schandry N, Arteaga-Salas JM, Israel L, Imhof A, Janzen CJ (2012) DOT1A-dependent H3K76 methylation is required for replication regulation in *Trypanosoma brucei*. *Nucleic Acids Res*, **40**: 10302–10311
- Geiduschek EP, Kassavetis GA (2001) The RNA polymerase III transcription apparatus. *J Mol Biol*, **310**: 1–26
- Gelfman S, Burstein D, Penn O, Savchenko A, Amit M, Schwartz S, Pupko T, Ast G (2012) Changes in exon-intron structure during vertebrate evolution affect the splicing pattern of exons. *Genome Res*, **22**: 35–50
- Geoffrey MC (2000) The Cell.
- Gershenzon NI, Ioshikhes IP (2005) Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics*, **21**: 1295–1300
- Ghosh A, Shuman S, Lima CD (2011) Structural insights to how mammalian capping enzyme reads the CTD code. *Mol Cell*, **43**: 299–310
- Gilchrist DA, Dos Santos G, Fargo DC, Xie B, Gao Y, Li L, Adelman K (2010) Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell*, **143**: 540–551
- 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
- Gilinger G, Bellofatto V (2001) Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. *Nucleic Acids Res*, **29**: 1556–1564
- Ginsburg DS, Anlembom TE, Wang J, Patel SR, Li B, Hinnebusch AG (2014) NuA4 links methylation of histone H3 lysines 4 and 36 to acetylation of histones H4 and H3. *J Biol Chem*, **289**: 32656–32670
- Goldmark JP, Fazio TG, Estep PW, Church GM, Tsukiyama T (2000) The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. *Cell*, **103**: 423–433
- Guth S, Tange TØ, Kellenberger E, Valcárcel J (2001) Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. *Mol Cell Biol*, **21**: 7673–7681
- Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **458**: 223–227
- Hampsey M, Reinberg D (2003) Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell*, **113**: 429–432
- Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**: 99–104
- Hargreaves DC, Horng T, Medzhitov R (2009) Control of inducible gene expression by signal-

- dependent transcriptional elongation. *Cell*, **138**: 129–145
- Harlow ED, Laboratory DL-...CSH, 1988 A laboratory manual. *genesdevcshlporg*,
- Hartley PD, Madhani HD (2009) Mechanisms that specify promoter nucleosome location and identity. *Cell*, **137**: 445–458
- He Y, Fang J, Taatjes DJ, Nogales E (2013) Structural visualization of key steps in human transcription initiation. *Nature*, **495**: 481–486
- Herbert A, Genetica AR, 1999 (1999) Left-handed Z-DNA: structure and function. *Springer*, **106**:
- Hernández-Rivas R, Martínez-Calvillo S, Romero M, Hernández R (1992) Trypanosoma cruzi 5S rRNA genes: molecular cloning, structure and chromosomal organization. *FEMS Microbiol Lett*, **71**: 63–67
- Hertz-Fowler C, Figueiredo LM, Quail MA, Becker M, Jackson A, Bason N, Brooks K, Churcher C, Fahkro S, Goodhead I, Heath P, Kartvelishvili M, Mungall K, Harris D, Hauser H, Sanders M, Saunders D, Seeger K, Sharp S, Taylor JE, Walker D, White B, Young R, Cross GA, Rudenko G, Barry JD, Louis EJ, Berriman M (2008) Telomeric expression sites are highly conserved in Trypanosoma brucei. *PLoS One*, **3**: e3527
- Hewish DR, Burgoyne LA (1973) Chromatin substructure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem Biophys Res Commun*, **52**: 504–510
- Hirschhorn JN, Brown SA, Clark CD, Winston F (1992) Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Genes Dev*, **6**: 2288–2298
- Hirumi H, Hirumi K (1994) Axenic culture of African trypanosome bloodstream forms. *Parasitol Today*, **10**: 80–84
- Hoek M, Engstler M, Cross GA (2000) Expression-site-associated gene 8 (ESAG8) of Trypanosoma brucei is apparently essential and accumulates in the nucleolus. *J Cell Sci*, **113**: 3959–3968
- Hoffman EA, Frey BL, Smith LM, Auble DT (2015) Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J Biol Chem*, **290**: 26404–26411
- Hu P, Wu S, Sun Y, Yuan CC, Kobayashi R, Myers MP, Hernandez N (2002) Characterization of human RNA polymerase III identifies orthologues for Saccharomyces cerevisiae RNA polymerase III subunits. *Mol Cell Biol*, **22**: 8044–8055
- Huang J, Van der Ploeg LH (1991) Requirement of a polypyrimidine tract for trans-splicing in trypanosomes: discriminating the PARP promoter from the immediately adjacent 3' splice acceptor site. *EMBO J*, **10**: 3877–3885
- Hughes AL, Jin Y, Rando OJ, Struhl K (2012) A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. *Mol Cell*, **48**: 5–15
- Hughes AL, Rando OJ (2014) Mechanisms underlying nucleosome positioning in vivo. *Annu Rev Biophys*, **43**: 41–63
- Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchetti G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RM, Cronin A, Cruz AK, Davies RM, De Gaudenzi J, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N, Frasch AC, Fraser A, Fuchs M, Gabel C, Goble A, Goffeau A, Harris D, Hertz-Fowler C, Hilbert H, Horn D, Huang Y, Klages S, Knights A, Kube M, Larke N, Litvin L, Lord A, Louie T, Marra M, Masuy D, Matthews K, Michaeli S, Mottram JC, Muller-Auer S, Munden H, Nelson S, Norbertczak H, Oliver K, O'neil S, Pentony M, Pohl TM, Price C, Purnelle B, Quail MA, Rabinowitsch E, Reinhardt R, Rieger M, Rinta J, Robben J, Robertson L, Ruiz JC, Rutter S, Saunders D, Schafer M, Schein J, Schwartz DC, Seeger K, Seyler A, Sharp S, Shin H, Sivam D, Squares R, Squares S, Tosato V, Vogt C, Volckaert G, Wambutt R, Warren T, Wedler H, Woodward J, Zhou S, Zimmermann W, Smith DF, Blackwell JM, Stuart KD, Barrell B, Myler PJ (2005) The genome of the kinetoplastid parasite, Leishmania major. *Science*, **309**: 436–442
- Janzen CJ, Fernandez JP, Deng H, Diaz R, Hake SB, Cross GA (2006a) Unusual histone modifications in Trypanosoma brucei. *FEBS Lett*, **580**: 2306–2310
- Janzen CJ, Hake SB, Lowell JE, Cross GA (2006b) Selective di- or trimethylation of histone H3 lysine 76 by two DOT1 homologs is important for cell cycle regulation in Trypanosoma brucei. *Mol Cell*, **23**: 497–507
- Jiang C, Pugh BF (2009) A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome. *Genome Biol*, **10**: R109
- Jin C, Felsenfeld G (2007) Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev*, **21**: 1519–1529
- Jocelyn EK, Benjamin L, Elliott SG, Stephen TK (2011) Lewin's Genes X.
- Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science*, **316**: 1497–1502
- Johnson PJ, Kooter JM, Borst P (1987) Inactivation of transcription by UV irradiation of T. brucei provides evidence for a multicistronic transcription unit including a VSG gene. *Cell*, **51**: 273–281
- Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*, **13**: 484–492
- Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT (2008) The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol*, **20**: 253–259
- Juven-Gershon T, Kadonaga JT (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol*, **339**: 225–229
- Kao CF, Hillyer C, Tsukuda T, Henry K, Berger S, Osley MA (2004) Rad6 plays a role in

- transcriptional activation through ubiquitylation of histone H2B. *Genes Dev*, **18**: 184–195
- Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, Segal E (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, **458**: 362–366
- Kawahara T, Siegel TN, Ingram AK, Alsford S, Cross GA, Horn D (2008) Two essential MYST-family proteins display distinct roles in histone H4K10 acetylation and telomeric silencing in trypanosomes. *Mol Microbiol*, **69**: 1054–1068
- Kelly TK, Miranda TB, Liang G, Berman BP, Lin JC, Tanay A, Jones PA (2010) H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. *Mol Cell*, **39**: 901–911
- Kephart DD, Marshall NF, Price DH (1992) Stability of *Drosophila* RNA polymerase II elongation complexes in vitro. *Mol Cell Biol*, **12**: 2067–2077
- Kielkopf CL, Rodionova NA, Green MR, Burley SK (2001) A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell*, **106**: 595–605
- Kim JL, Nikolov DB, Burley SK (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**: 520–527
- Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B (2005) A high-resolution map of active promoters in the human genome. *Nature*, **436**: 876–880
- Kimura A, Umehara T, Horikoshi M (2002) Chromosomal gradient of histone acetylation established by Sas2p and Sir2p functions as a shield against gene silencing. *Nat Genet*, **32**: 370–377
- Kireeva ML, Walter W, Tchernajenko V, Bondarenko V, Kashlev M, Studitsky VM (2002) Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. *Mol Cell*, **9**: 541–552
- Kobor MS, Venkatasubrahmanyam S, Meneghini MD, Gin JW, Jennings JL, Link AJ, Madhani HD, Rine J (2004) A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. *PLoS Biol*, **2**: E131
- Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, Spicuglia S, de la Chapelle AL, Heidemann M, Hintermair C, Eick D, Gut I, Ferrier P, Andrau JC (2011) Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol*, **18**: 956–963
- Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C (2010) The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLoS Pathog*, **6**: e1001090
- Kooter JM, Borst P (1984) Alpha-amanitin-insensitive transcription of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. *Nucleic Acids Res*, **12**: 9457–9472
- Korber P, Hörz W (2004) In vitro assembly of the characteristic chromatin organization at the yeast PHO5 promoter by a replication-independent extract system. *J Biol Chem*, **279**: 35113–35120
- Kornberg RD, Stryer L (1988) Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res*, **16**: 6677–6690
- Kornberg RD, Thomas JO (1974) Chromatin structure; oligomers of the histones. *Science*, **184**: 865–868
- Kulaeva OI, Gaykalova DA, Pestov NA, Golovastov VV, Vassilyev DG, Artsimovitch I, Studitsky VM (2009) Mechanism of chromatin remodeling and recovery during passage of RNA polymerase II. *Nat Struct Mol Biol*, **16**: 1272–1278
- Kuryan BG, Kim J, Tran NN, Lombardo SR, Venkatesh S, Workman JL, Carey M (2012) Histone density is maintained during transcription mediated by the chromatin remodeler RSC and histone chaperone NAP1 in vitro. *Proc Natl Acad Sci U S A*, **109**: 1931–1936
- Lafer EM, Möller A, Nordheim A, Stollar BD, Rich A (1981) Antibodies specific for left-handed Z-DNA. *Proc Natl Acad Sci U S A*, **78**: 3546–3550
- Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH (1998) New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev*, **12**: 34–44
- Landolin JM, Johnson DS, Trinklein ND, Aldred SF, Medina C, Shulha H, Weng Z, Myers RM (2010) Sequence features that drive human promoter function and tissue specificity. *Genome Res*, **20**: 890–898
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**: 357–359
- Lantermann A, Strålfors A, Fagerström-Billai F, Korber P, Ekwall K (2009) Genome-wide mapping of nucleosome positions in *Schizosaccharomyces pombe*. *Methods*, **48**: 218–225
- Lantermann AB, Straub T, Strålfors A, Yuan GC, Ekwall K, Korber P (2010) *Schizosaccharomyces pombe* genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of *Saccharomyces cerevisiae*. *Nat Struct Mol Biol*, **17**: 251–257
- Larabee RN, Krogan NJ, Xiao T, Shibata Y, Hughes TR, Greenblatt JF, Strahl BD (2005) BUR kinase selectively regulates H3 K4 trimethylation and H2B ubiquitylation through recruitment of the PAF elongation complex. *Curr Biol*, **15**: 1487–1493
- Lecordier L, Devaux S, Uzureau P, Dierick JF, Walgraffe D, Poelvoorde P, Pays E, Vanhamme L (2007) Characterization of a TFIIF homologue from *Trypanosoma brucei*. *Mol Microbiol*, **64**: 1164–1181
- Lee C, Li X, Hechmer A, Eisen M, Biggin MD, Venters BJ, Jiang C, Li J, Pugh BF, Gilmour DS (2008) NELF and GAGA factor are linked to promoter-proximal pausing at many genes in *Drosophila*. *Mol Cell Biol*, **28**: 3290–3300
- Lee JH, Nguyen TN, Schimanski B, Günzl A (2007a) Spliced leader RNA gene transcription in

- Trypanosoma brucei requires transcription factor TFIIH. *Eukaryot Cell*, **6**: 641–649
- Lee K, Kim SC, Jung I, Kim K, Seo J, Lee HS, Bogu GK, Kim D, Lee S, Lee B, Choi JK (2013) Genetic landscape of open chromatin in yeast. *PLoS Genet*, **9**: e1003229
- Lee MG (1996) An RNA polymerase II promoter in the hsp70 locus of Trypanosoma brucei. *Mol Cell Biol*, **16**: 1220–1230
- Lee TI, Johnstone SE, Young RA (2006) Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc*, **1**: 729–748
- Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C (2007b) A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet*, **39**: 1235–1244
- Lenhard B, Sandelin A, Carninci P (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet*, **13**: 233–245
- Lewis BA, Kim TK, Orkin SH (2000) A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. *Proc Natl Acad Sci U S A*, **97**: 7172–7177
- Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, **26**: 589–595
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 GPDPs (2009a) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**: 2078–2079
- Li H, Xiao J, Li J, Lu L, Feng S, Dröge P (2009b) Human genomic Z-DNA segments probed by the Z alpha domain of ADAR1. *Nucleic Acids Res*, **37**: 2737–2746
- Li Z, Schug J, Tuteja G, White P, Kaestner KH (2011) The nucleosome map of the mammalian liver. *Nat Struct Mol Biol*, **18**: 742–746
- Liang XH, Haritan A, Uliel S, Michaeli S (2003) trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot Cell*, **2**: 830–840
- 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
- Lieleg C, Krietenstein N, Walker M, Korber P (2014) Nucleosome positioning in yeasts: methods, maps, and mechanisms. *Chromosoma*
- Liu B, Molina H, Kalume D, Pandey A, Griffith JD, Englund PT (2006) Role of p38 in replication of Trypanosoma brucei kinetoplast DNA. *Mol Cell Biol*, **26**: 5382–5393
- Liu LF, Wang JC (1987) Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A*, **84**: 7024–7027
- Liu R, Liu H, Chen X, Kirby M, Brown PO, Zhao K (2001) Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. *Cell*, **106**: 309–318
- Lowary PT, Widom J (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol*, **276**: 19–42
- Lowell JE, Cross GA (2004) A variant histone H3 is enriched at telomeres in Trypanosoma brucei. *J Cell Sci*, **117**: 5937–5947
- Lowell JE, Kaiser F, Janzen CJ, Cross GA (2005) Histone H2AZ dimerizes with a novel variant H2B and is enriched at repetitive DNA in Trypanosoma brucei. *J Cell Sci*, **118**: 5721–5730
- Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**: 251–260
- Luse DS (2012) Rethinking the role of TFIIF in transcript initiation by RNA polymerase II. *Transcription*, **3**: 156–159
- Luse DS (2014) The RNA polymerase II preinitiation complex. Through what pathway is the complex assembled. *Transcription*, **5**: e27050
- Madhani HD, Guthrie C (1994) Dynamic RNA-RNA interactions in the spliceosome. *Annu Rev Genet*, **28**: 1–26
- Mahony S, Pugh BF (2015) Protein-DNA binding in high-resolution. *Crit Rev Biochem Mol Biol*, **50**: 269–283
- Mair G, Shi H, Li H, Djikeng A, Aviles HO, Rna JRB, 2000 A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *cambridgeorg*,
- Malik HS, Henikoff S (2003) Phylogenomics of the nucleosome. *Nat Struct Biol*, **10**: 882–891
- Malvy D, Chappuis F (2011) Sleeping sickness. *Clin Microbiol Infect*, **17**: 986–995
- Mandava V, Fernandez JP, Deng H, Janzen CJ, Hake SB, Cross GA (2007) Histone modifications in Trypanosoma brucei. *Mol Biochem Parasitol*, **156**: 41–50
- Marshall NF, Price DH (1992) Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol Cell Biol*, **12**: 2078–2090
- 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
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*,
- Martínez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ (2003) Transcription of Leishmania major Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell*, **11**: 1291–1299
- Maruyama A, Mimura J, Harada N, Itoh K (2013) Nrf2 activation is associated with Z-DNA formation in the human HO-1 promoter. *Nucleic Acids Res*, **41**: 5223–5234
- Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC, Gilmour DS, Albert I, Pugh BF (2008) Nucleosome organization in the Drosophila genome. *Nature*, **453**: 358–362

- McAndrew M, Graham S, Hartmann C, Clayton C (1998) Testing promoter activity in the trypanosome genome: isolation of a metacyclic-type VSG promoter, and unexpected insights into RNA polymerase II transcription. *Exp Parasitol*, **90**: 65–76
- McCall M, Brown T, Kennard O (1985) The crystal structure of d(G-G-G-G-C-C-C). A model for poly(dG).poly(dC). *J Mol Biol*, **183**: 385–396
- Melville SE, Leech V, Navarro M, Cross GA (2000) The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* stock 427. *Mol Biochem Parasitol*, **111**: 261–273
- Millar CB, Xu F, Zhang K, Grunstein M (2006) Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast. *Genes Dev*, **20**: 711–722
- Min IM, Waterfall JJ, Core LJ, Munroe RJ, Schimenti J, Lis JT (2011) Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev*, **25**: 742–754
- Mizuguchi G, Shen X, Landry J, Wu WH, Sen S, Wu C (2004) ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science*, **303**: 343–348
- Möbius W, Gerland U (2010) Quantitative test of the barrier nucleosome model for statistical positioning of nucleosomes up- and downstream of transcription start sites. *PLoS Comput Biol*, **6**: e1000807
- Mohrmann L, Verrijzer CP (2005) Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. *Biochim Biophys Acta*, **1681**: 59–73
- Möller A, Gabriels JE, Lafer EM, Nordheim A, Rich A, Stollar BD (1982) Monoclonal antibodies recognize different parts of Z-DNA. *J Biol Chem*, **257**: 12081–12085
- Morse RH (2007) Transcription factor access to promoter elements. *J Cell Biochem*, **102**: 560–570
- Murakami K, Calero G, Brown CR, Liu X, Davis RE, Boeger H, Kornberg RD (2013) Formation and fate of a complete 31-protein RNA polymerase II transcription preinitiation complex. *J Biol Chem*, **288**: 6325–6332
- 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
- Naftelberg S, Schor IE, Ast G, Kornblihtt AR (2015) Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu Rev Biochem*, **84**: 165–198
- Nahkuri S, Taft RJ, Mattick JS (2009) Nucleosomes are preferentially positioned at exons in somatic and sperm cells. *Cell Cycle*, **8**: 3420–3424
- Nakanishi S, Sanderson BW, Delventhal KM, Bradford WD, Staehling-Hampton K, Shilatifard A (2008) A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation. *Nat Struct Mol Biol*, **15**: 881–888
- Navarro M, Gull K (2001) A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*. *Nature*, **414**: 759–763
- 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
- Nekrasov M, Amrichova J, Parker BJ, Soboleva TA, Jack C, Williams R, Huttley GA, Tremethick DJ (2012) Histone H2A.Z inheritance during the cell cycle and its impact on promoter organization and dynamics. *Nat Struct Mol Biol*, **19**: 1076–1083
- Nelson HC, Finch JT, Luisi BF, Klug A (1987) The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature*, **330**: 221–226
- Ng HH, Ciccone DN, Morshead KB, Oettinger MA, Struhl K (2003a) Lysine-79 of histone H3 is hypomethylated at silenced loci in yeast and mammalian cells: a potential mechanism for position-effect variegation. *Proc Natl Acad Sci U S A*, **100**: 1820–1825
- Ng HH, Dole S, Struhl K (2003b) The Rtf1 component of the Paf1 transcriptional elongation complex is required for ubiquitination of histone H2B. *J Biol Chem*, **278**: 33625–33628
- Ng HH, Robert F, Young RA, Struhl K (2003c) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell*, **11**: 709–719
- Nguyen TN, Müller LS, Park SH, Siegel TN, Günzl A (2014) Promoter occupancy of the basal class I transcription factor A differs strongly between active and silent VSG expression sites in *Trypanosoma brucei*. *Nucleic Acids Res*, **42**: 3164–3176
- Ni T, Corcoran DL, Rach EA, Song S, Spana EP, Gao Y, Ohler U, Zhu J (2010) A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat Methods*, **7**: 521–527
- Nilsson D, Gunasekera K, Mani J, Osteras M, Farinelli L, Baerlocher L, Roditi I, Ochsenreiter T (2010) Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. *PLoS Pathog*, **6**: e1001037
- Nishida H, Motoyama T, Yamamoto S, Aburatani H, Osada H (2009) Genome-wide maps of mono- and di-nucleosomes of *Aspergillus fumigatus*. *Bioinformatics*, **25**: 2295–2297
- Olins AL, Olins DE (1974) Spheroid chromatin units (v bodies). *Science*, **183**: 330–332
- Olson WK, Zhurkin VB (2011) Working the kinks out of nucleosomal DNA. *Curr Opin Struct Biol*, **21**: 348–357
- Ozonov EA, van Nimwegen E (2013) Nucleosome free regions in yeast promoters result from competitive binding of transcription factors that interact with chromatin modifiers. *PLoS Comput Biol*, **9**: e1003181
- Palenchar JB, Liu W, Palenchar PM, Bellofatto V (2006) A divergent transcription factor TFIIIB in trypanosomes is required for RNA polymerase II-dependent spliced leader RNA transcription and cell viability. *Eukaryot Cell*, **5**: 293–300

- Papamichos-Chronakis M, Watanabe S, Rando OJ, Peterson CL (2011) Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell*, **144**: 200–213
- Pavri R, Zhu B, Li G, Trojer P, Mandal S, Shilatifard A, Reinberg D (2006) Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell*, **125**: 703–717
- Perales R, Zhang L, Bentley D (2011) Histone occupancy in vivo at the 601 nucleosome binding element is determined by transcriptional history. *Mol Cell Biol*, **31**: 3485–3496
- Perry KL, Watkins KP, Agabian N (1987) Trypanosome mRNAs have unusual “cap 4” structures acquired by addition of a spliced leader. *Proc Natl Acad Sci U S A*, **84**: 8190–8194
- Peterlin BM, Price DH (2006) Controlling the elongation phase of transcription with P-TEFb. *Mol Cell*, **23**: 297–305
- Pray-Grant MG, Daniel JA, Schieltz D, Yates JR, Grant PA (2005) Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature*, **433**: 434–438
- Prunell A, Kornberg RD (1982) Variable center to center distance of nucleosomes in chromatin. *J Mol Biol*, **154**: 515–523
- Quintales L, Vázquez E, Antequera F (2015) Comparative analysis of methods for genome-wide nucleosome cartography. *Brief Bioinform*, **16**: 576–587
- Rach EA, Winter DR, Benjamin AM, Corcoran DL, Ni T, Zhu J, Ohler U (2011) Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. *PLoS Genet*, **7**: e1001274
- Radman-Livaja M, Rando OJ (2010) Nucleosome positioning: how is it established, and why does it matter. *Dev Biol*, **339**: 258–266
- 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
- Rahmouni AR, Wells RD (1989) Stabilization of Z DNA in vivo by localized supercoiling. *Science*, **246**: 358–363
- Raisner RM, Hartley PD, Meneghini MD, Bao MZ, Liu CL, Schreiber SL, Rando OJ, Madhani HD (2005) Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell*, **123**: 233–248
- Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, Doty KR, Black JC, Hoffmann A, Carey M, Smale ST (2009) A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell*, **138**: 114–128
- Rhee HS, Pugh BF (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*, **147**: 1408–1419
- Rhee HS, Pugh BF (2012) Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*, **483**: 295–301
- Rich A, Zhang S (2003) Timeline: Z-DNA: the long road to biological function. *Nat Rev Genet*, **4**: 566–572
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. *Nat Biotechnol*, **29**: 24–26
- Robzyk K, Recht J, Osley MA (2000) Rad6-dependent ubiquitination of histone H2B in yeast. *Science*, **287**: 501–504
- Roeder RG (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci*, **21**: 327–335
- Roy AL, Singer DS (2015) Core promoters in transcription: old problem, new insights. *Trends Biochem Sci*, **40**: 165–171
- Ruan JP, Arhin GK, Ullu E, Tschudi C (2004) Functional characterization of a Trypanosoma brucei TATA-binding protein-related factor points to a universal regulator of transcription in trypanosomes. *Mol Cell Biol*, **24**: 9610–9618
- Ruskin B, Krainer AR, Maniatis T, Green MR (1984) Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell*, **38**: 317–331
- Saha A, Wittmeyer J, Cairns BR (2006) Chromatin remodelling: the industrial revolution of DNA around histones. *Nat Rev Mol Cell Biol*, **7**: 437–447
- Sainsbury S, Niesser J, Cramer P (2013) Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature*, **493**: 437–440
- Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume DA (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet*, **8**: 424–436
- Santos-Rosa H, Bannister AJ, Dehe PM, Géli V, Kouzarides T (2004) Methylation of H3 lysine 4 at euchromatin promotes Sir3p association with heterochromatin. *J Biol Chem*, **279**: 47506–47512
- Satchwell SC, Drew HR, Travers AA (1986) Sequence periodicities in chicken nucleosome core DNA. *J Mol Biol*, **191**: 659–675
- Saxonov S, Berg P, Brutlag DL (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A*, **103**: 1412–1417
- Scahill MD, Pastar I, Cross GA (2008) CRE recombinase-based positive-negative selection systems for genetic manipulation in Trypanosoma brucei. *Mol Biochem Parasitol*, **157**: 73–82
- Schimanski B, Brandenburg J, Nguyen TN, Caimano MJ, Günzl A (2006) A TFIIB-like protein is indispensable for spliced leader RNA gene transcription in Trypanosoma brucei. *Nucleic Acids Res*, **34**: 1676–1684
- Schimanski B, Nguyen TN, Günzl A (2005) Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA

- gene transcription in *Trypanosoma brucei*. *Mol Cell Biol*, **25**: 7303–7313
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods*, **9**: 676–682
- Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell*, **132**: 887–898
- Schroth GP, Chou PJ, Ho PS (1992) Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. *J Biol Chem*, **267**: 11846–11855
- Schultz LD, Hall BD (1976) Transcription in yeast: alpha-amanitin sensitivity and other properties which distinguish between RNA polymerases I and III. *Proc Natl Acad Sci U S A*, **73**: 1029–1033
- Schumann Burkard G, Jutzi P, Roditi I (2011) Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters. *Mol Biochem Parasitol*, **175**: 91–94
- Schwartz S, Ast G (2010) Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. *EMBO J*, **29**: 1629–1636
- Schwartz S, Meshorer E, Ast G (2009) Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol*, **16**: 990–995
- Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, Wang JP, Widom J (2006) A genomic code for nucleosome positioning. *Nature*, **442**: 772–778
- Segal E, Widom J (2009) Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr Opin Struct Biol*, **19**: 65–71
- Shin SI, Ham S, Park J, Seo SH, Lim CH, Jeon H, Huh J, Roh TY (2016) Z-DNA-forming sites identified by ChIP-Seq are associated with actively transcribed regions in the human genome. *DNA Res*
- Siegel TN, Gunasekera K, Cross GA, Ochsenreiter T (2011) Gene expression in *Trypanosoma brucei*: lessons from high-throughput RNA sequencing. *Trends Parasitol*, **27**: 434–441
- Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, Fenyo D, Wang X, Dewell S, Cross GA (2009) Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev*, **23**: 1063–1076
- Siegel TN, Hekstra DR, Wang X, Dewell S, Cross GA (2010) Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. *Nucleic Acids Res*, **38**: 4946–4957
- Siegel TN, Kawahara T, Degrasse JA, Janzen CJ, Horn D, Cross GA (2008) Acetylation of histone H4K4 is cell cycle regulated and mediated by HAT3 in *Trypanosoma brucei*. *Mol Microbiol*, **67**: 762–771
- Siegel TN, Tan KS, Cross GA (2005) Systematic study of sequence motifs for RNA trans splicing in *Trypanosoma brucei*. *Mol Cell Biol*, **25**: 9586–9594
- Sims RJ, Millhouse S, Chen CF, Lewis BA, Erdjument-Bromage H, Tempst P, Manley JL, Reinberg D (2007) Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol Cell*, **28**: 665–676
- Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. *Annu Rev Biochem*, **72**: 449–479
- Small EC, Xi L, Wang JP, Widom J, Licht JD (2014) Single-cell nucleosome mapping reveals the molecular basis of gene expression heterogeneity. *Proc Natl Acad Sci U S A*, **111**: E2462–71
- Smith JL, Levin JR, Ingles CJ, Agabian N (1989) In trypanosomes the homolog of the largest subunit of RNA polymerase II is encoded by two genes and has a highly unusual C-terminal domain structure. *Cell*, **56**: 815–827
- Smolle M, Workman JL (2013) Transcription-associated histone modifications and cryptic transcription. *Biochim Biophys Acta*, **1829**: 84–97
- Staley JP, Guthrie C (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, **92**: 315–326
- Steger DJ, Lefterova MI, Ying L, Stonestrom AJ, Schupp M, Zhuo D, Vakoc AL, Kim JE, Chen J, Lazar MA, Blobel GA, Vakoc CR (2008) DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol Cell Biol*, **28**: 2825–2839
- Struhl K, Segal E (2013) Determinants of nucleosome positioning. *Nat Struct Mol Biol*, **20**: 267–273
- Suka N, Luo K, Grunstein M (2002) Sir2p and Sas2p opposingly regulate acetylation of yeast histone H4 lysine16 and spreading of heterochromatin. *Nat Genet*, **32**: 378–383
- Sun ZW, Allis CD (2002) Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature*, **418**: 104–108
- Suter B, Schnappauf G, Thoma F (2000) Poly(dA:dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. *Nucleic Acids Res*, **28**: 4083–4089
- Suto RK, Clarkson MJ, Tremethick DJ, Luger K (2000) Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nat Struct Biol*, **7**: 1121–1124
- Talbert PB, Henikoff S (2010) Histone variants—ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol*, **11**: 264–275
- Teif VB, Vainshtein Y, Caudron-Herger M, Mallm JP, Marth C, Höfer T, Rippe K (2012) Genome-wide nucleosome positioning during embryonic stem cell development. *Nat Struct Mol Biol*, **19**: 1185–1192
- Thatcher TH, Gorovsky MA (1994) Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. *Nucleic Acids Res*, **22**: 174–179

- Tilgner H, Nikolaou C, Althammer S, Sammeth M, Beato M, Valcarcel J, Guigo R (2009) Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol*, **16**: 996–1001
- Tillo D, Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Field Y, Lieb JD, Widom J, Segal E, Hughes TR (2010) High nucleosome occupancy is encoded at human regulatory sequences. *PLoS One*, **5**: e9129
- Tirode F, Busso D, Coin F, Egly JM (1999) Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol Cell*, **3**: 87–95
- Tirosh I, Barkai N (2008) Two strategies for gene regulation by promoter nucleosomes. *Genome Res*, **18**: 1084–1091
- Tolstorukov MY, Volfovsky N, Stephens RM, Park PJ (2011) Impact of chromatin structure on sequence variability in the human genome. *Nat Struct Mol Biol*, **18**: 510–515
- Tran HG, Steger DJ, Iyer VR, Johnson AD (2000) The chromo domain protein chd1p from budding yeast is an ATP-dependent chromatin-modifying factor. *EMBO J*, **19**: 2323–2331
- Tsankov A, Yanagisawa Y, Rhind N, Regev A, Rando OJ (2011) Evolutionary divergence of intrinsic and trans-regulated nucleosome positioning sequences reveals plastic rules for chromatin organization. *Genome Res*, **21**: 1851–1862
- Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ (2010) The role of nucleosome positioning in the evolution of gene regulation. *PLoS Biol*, **8**: e1000414
- Urwyler S, Studer E, Renggli CK, Roditi I (2007) A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of *Trypanosoma brucei*. *Mol Microbiol*, **63**: 218–228
- Uzureau P, Daniels JP, Walgraffe D, Wickstead B, Pays E, Gull K, Vanhamme L (2008) Identification and characterization of two trypanosome TFIIIS proteins exhibiting particular domain architectures and differential nuclear localizations. *Mol Microbiol*, **69**: 1121–1136
- Valay JG, Simon M, Dubois MF, Bensaude O, Facca C, Faye G (1995) The KIN28 gene is required both for RNA polymerase II mediated transcription and phosphorylation of the Rpb1p CTD. *J Mol Biol*, **249**: 535–544
- Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K, Sidow A, Fire A, Johnson SM (2008) A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res*, **18**: 1051–1063
- Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A (2011) Determinants of nucleosome organization in primary human cells. *Nature*, **474**: 516–520
- van Heeringen SJ, Akhtar W, Jacobi UG, Akkers RC, Suzuki Y, Veenstra GJ (2011) Nucleotide composition-linked divergence of vertebrate core promoter architecture. *Genome Res*, **21**: 410–421
- van Leeuwen F, Gafken PR, Gottschling DE (2002) Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell*, **109**: 745–756
- Vasquez JJ, Hon CC, Vanselow JT, Schlosser A, Siegel TN (2014) Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages. *Nucleic Acids Res*, **42**: 3623–3637
- Venkatesh S, Workman JL (2015) Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol*,
- Venters BJ, Pugh BF (2009a) How eukaryotic genes are transcribed. *Crit Rev Biochem Mol Biol*, **44**: 117–141
- Venters BJ, Pugh BF (2009b) A canonical promoter organization of the transcription machinery and its regulators in the *Saccharomyces* genome. *Genome Res*, **19**: 360–371
- Vickerman K (1969) On the surface coat and flagellar adhesion in trypanosomes. *J Cell Sci*, **5**: 163–193
- 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
- Wal M, Pugh BF (2012) Genome-wide mapping of nucleosome positions in yeast using high-resolution MNase ChIP-Seq. *Methods Enzymol*, **513**: 233–250
- Wang AH, Quigley GJ, Kolpak FJ, Crawford JL, van Boom JH, van der Marel G, Rich A (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*, **282**: 680–686
- Wang G, Vasquez KM (2007) Z-DNA, an active element in the genome. *Front Biosci*, **12**: 4424–4438
- Wedel C, Förstner KU, Derr R, Siegel TN (2017) GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. *EMBO J*,
- Wedel C, Siegel TN (2017) Genome-wide analysis of chromatin structures in *Trypanosoma brucei* using high-resolution MNase-ChIP-seq. *Exp Parasitol*,
- Weiner A, Hughes A, Yassour M, Rando OJ, Friedman N (2010) High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome Res*, **20**: 90–100
- Westenberger SJ, Cui L, Dharia N, Winzeler E, Cui L (2009) Genome-wide nucleosome mapping of *Plasmodium falciparum* reveals histone-rich coding and histone-poor intergenic regions and chromatin remodeling of core and subtelomeric genes. *BMC Genomics*, **10**: 610
- Widom J (1989) Toward a unified model of chromatin folding. *Annu Rev Biophys Chem*, **18**: 365–395
- Widom J (1998) Structure, dynamics, and function of chromatin in vitro. *Annu Rev Biophys Biomol Struct*, **27**: 285–327

- Willis IM (1993) RNA polymerase III. Genes, factors and transcriptional specificity. *Eur J Biochem*, **212**: 1–11
- Wippo CJ, Israel L, Watanabe S, Hochheimer A, Peterson CL, Korber P (2011) The RSC chromatin remodelling enzyme has a unique role in directing the accurate positioning of nucleosomes. *EMBO J*, **30**: 1277–1288
- Wirtz E, Clayton C (1995) Inducible gene expression in trypanosomes mediated by a prokaryotic repressor. *Science*, **268**: 1179–1183
- Wirtz E, Leal S, Ochatt C, Cross GA (1999) A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol Biochem Parasitol*, **99**: 89–101
- Wittig B, Wöfl S, Dorbic T, Vahrson W, Rich A (1992) Transcription of human c-myc in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. *EMBO J*, **11**: 4653–4663
- Wong B, Chen S, Kwon JA, Rich A (2007) Characterization of Z-DNA as a nucleosome-boundary element in yeast *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*, **104**: 2229–2234
- Wood A, Schneider J, Dover J, Johnston M, Shilatifard A (2003) The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *J Biol Chem*, **278**: 34739–34742
- Wood A, Schneider J, Dover J, Johnston M, Shilatifard A (2005) The Bur1/Bur2 complex is required for histone H2B monoubiquitination by Rad6/Bre1 and histone methylation by COMPASS. *Mol Cell*, **20**: 589–599
- Wright JR, Siegel TN, Cross GA (2010) Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in *Trypanosoma brucei*. *Mol Biochem Parasitol*, **172**: 141–144
- Wu S, Romfo CM, Nilsen TW, Green MR (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature*, **402**: 832–835
- Wu WH, Alami S, Luk E, Wu CH, Sen S, Mizuguchi G, Wei D, Wu C (2005) Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange. *Nat Struct Mol Biol*, **12**: 1064–1071
- Yamamoto YY, Yoshitsugu T, Sakurai T, Seki M, Shinozaki K, Obokata J (2009) Heterogeneity of Arabidopsis core promoters revealed by high-density TSS analysis. *Plant J*, **60**: 350–362
- Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E (2007) Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene*, **389**: 52–65
- Yazdi PG, Pedersen BA, Taylor JF, Khattab OS, Chen YH, Chen Y, Jacobsen SE, Wang PH (2015) Nucleosome Organization in Human Embryonic Stem Cells. *PLoS One*, **10**: e0136314
- Yu L, Morse RH (1999) Chromatin opening and transactivator potentiation by RAP1 in *Saccharomyces cerevisiae*. *Mol Cell Biol*, **19**: 5279–5288
- Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ (2005) Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science*, **309**: 626–630
- Zamore PD, Patton JG, Green MR (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature*, **355**: 609–614
- 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
- Zhang H, Roberts DN, Cairns BR (2005) Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell*, **123**: 219–231
- Zhang Y, Moqtaderi Z, Rattner BP, Euskirchen G, Snyder M, Kadonaga JT, Liu XS, Struhl K (2009) Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat Struct Mol Biol*, **16**: 847–852
- Zhang Z, Dietrich FS (2005) Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5' SAGE. *Nucleic Acids Res*, **33**: 2838–2851
- Zhang Z, Wippo CJ, Wal M, Ward E, Korber P, Pugh BF (2011) A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science*, **332**: 977–980
- Zlatanova J, Thakar A (2008) H2A.Z: view from the top. *Structure*, **16**: 166–179

Curriculum vitae

List of publications

Publications containing parts of this thesis:

Wedel C, Förstner KU, Derr R, Siegel TN (2017) GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. *EMBO J*, 36: 2581–2594

Wedel C, Siegel TN (2017) Genome-wide analysis of chromatin structures in *Trypanosoma brucei* using high-resolution MNase-ChIP-seq. *Exp Parasitol*, 180: 2–12

Other publications:

Vasquez JJ*, **Wedel C***, Cosentino RO, Siegel TN (2018) Exploiting CRISPR-Cas9 technology to investigate individual histone modifications. *Nucleic Acids Res*

Müller LSM, Cosentino RO, Förstner KU, Guizetti J, **Wedel C**, Kaplan N, Janzen CJ, Arampatzi P, Vogel J, Steinbiss S, Otto TD, Saliba AE, Sebra RP, Siegel TN (accepted) Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature*

* co-first authorship

Attended conferences and courses

Conferences

- 2017 7th Kinetoplastid Molecular Cell Biology Meeting 2017, Woods Hole, Massachusetts, USA, Poster: 'Exploiting CRISPR-Cas9 technology to investigate individual histone modifications'
- 2016 BSP Trypanosomiasis and Leishmaniasis Seminar 2016, České Budějovice, Czech Republic, Talk: 'GT-rich promoters drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes'
- 2016 Annual Meeting of the German Society for Parasitology 2016, Göttingen, Germany, Talk: 'Nucleosome positioning and DNA sequence-mediated RNA polymerase II transcription initiation in *Trypanosoma brucei*'
- 2015 Eureka 10th International GSLS Symposium 2015, Würzburg, Germany, Poster: 'Nucleosome positioning and nucleosome depleted regions in *Trypanosoma brucei*'
- 2015 6th Kinetoplastid Molecular Cell Biology Meeting 2015, Woods Hole, Massachusetts, USA, Poster: 'Does the DNA sequence affect nucleosome positioning in *T. brucei*'
- 2014 3rd Mol Micro Meeting 2014, Würzburg, Germany, Poster: 'Transcription initiation in *Trypanosoma brucei*. Does the DNA sequence matter?'

Transferable skills training

- 2015 – 2016 Mentoring Life Sciences Program, GSLS
- Skill Profiling and Self-Promotion
 - Tools and Strategies for a Successful Time- and Self-Management
 - Small Talk at International Conferences
 - Voice Box Training for Enhanced Vocal Performances
 - Conflict Management
 - Dos and Taboos of Business and Table Etiquette

GSLS workshops

- Software Carpentry for beginners, Dr. Konrad Förstner, Malvika Sharan, Markus Ankenbrand
- Introduction to biotech industries, Dr. Christian Grote-Westrick
- Cover letter & CV, Robert Zaal
- Job Interview Training, Robert Zaal
- Good Scientific Practice, Dr. Stephan Schröder-Köhne
- Self-awareness coaching, Cornelia C. Fink
- Scientific writing for PhD students, Dr. Andrew Davis
- Poster Design, Barry Drees

other workshops

- Vektorgrafiken erstellen mit Adobe Illustrator, Rechenzentrum Uni Würzburg

