Characterizing Long Non-Coding RNA at the DBF4 Gene Locus

By

Shawn Hughes

Thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science (MSc.) in Biology

The Faculty of Graduate Studies
Laurentian University
Sudbury, Ontario

©Shawn Hughes, 2015

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne
Faculty of Graduate Studies/Faculté des études supérieures

Title of Thesis
Titre de la thèse Characterizing Long Non-Coding RNA at the DBF4 Gene Locus

Name of Candidate
Nom du candidat Hughes, Shawn

Degree
Diplôme Master of Science

Department/Program
Département/Programme Biology

Date of Defence
Date de la soutenance April 26, 2016

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. Hoyun Lee
(Supervisor/Directeur(trice) de thèse)

Dr. Robert Lafrenie
(Committee member/Membre du comité)

Dr. Kabwe Nkongolo
(Committee member/Membre du comité)

(Committee member/Membre du comité)

Dr. Chris J. Brandl
(External Examiner/Examinateur externe)

(Internal Examiner/Examinateur interne)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Dr. David Lesbarrères
Monsieur David Lesbarrères
Dean, Faculty of Graduate Studies
Doyen, Faculté des études supérieures

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

I, **Shawn Hughes**, hereby grant to Laurentian University and/or its agents the non-exclusive license to archive and make accessible my thesis, dissertation, or project report in whole or in part in all forms of media, now or for the duration of my copyright ownership. I retain all other ownership rights to the copyright of the thesis, dissertation or project report. I also reserve the right to use in future works (such as articles or books) all or part of this thesis, dissertation, or project report. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that this copy is being made available in this form by the authority of the copyright owner solely for the purpose of private study and research and may not be copied or reproduced except as permitted by the copyright laws without written authority from the copyright owner.

ABSTRACT

Proper regulation of the cell cycle is essential for genetic stability and function. Mutations resulting in deregulation of the cell cycle are attributable to the development of diseases such as cancer. A large portion of the proteins responsible for regulating the cell cycle are found at DNA replication origins making replication origins an attractive target of research. Our lab has previously characterized a replication origin found in the promoter region of the human DBF4 locus which contains two zones at which two Origin Recognition Complexes bind and initiate replication in opposite directions in a specific manner termed Asymmetrical Bidirectional Replication. Within these same zones, our lab has also discovered the presence of transcription start sites for long non-coding RNA (lncRNA) transcripts. We've found that these non-coding transcripts are differentially transcribed from their neighboring genes and are resistant to RNA polymerase I, II and III inhibition. Results from these studies unexpectedly indicate the possibility of the mitochondrial RNA polymerase acting in the nucleus to transcribe the lncRNAs at the *DBF4* locus. These transcripts are also cyclically more stably expressed than the *DBF4* gene while retaining a relatively short half-life.

In an attempt to achieve a broader understanding of the replication process, the work in this thesis presents an effort to characterize and identify these lncRNAs initiating in the *DBF4* promoter region and provide new insights into their behavior.

AKNOWLEDGMENTS

First and foremost, I would like to give a very big thank you to my MSc. supervisor Dr. Hoyun Lee. I`ve been blessed with the opportunity to learn a lot through his insight and through this research opportunity. I would especially like to thank him for his patience and understanding during the time it took to write this thesis and for always keeping an open door in his office.

I would also like to thank my committee member and former lab director Dr. Robert Lafrenie for his guidance in the lab and for training me to use equipment such as the Flow Cytometry machine. I would also like to thank Dr. Kabwe Nkongolo for taking the time to go over my committee meeting presentations and provide his valuable input on my project.

I would like to thank Dr. Julia Romero whom I've never met but whose work I had the privilege of continuing. I would like to give a particularly warm thank you to Dr. James Knockleby for all the support in troubleshooting protocols and the intensive discussions on my research project. I would also like to thank Kathleen Kylie and Alison Douglas for helping me get started in the lab and showing me the ropes and a big thank you to all of the other Lee lab members who I've had the fortune to share lab space and equipment with. Lastly I would like to thank Dr. Sheetal Pundir for and stern (but welcome) motivational speeches and for pushing me to move forward.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABR	assymetrical bidirectional replication
ALDH8	aldehyde dehydrogenase 8 gene
Cdc6	cell division cycle 6
Cdc7	cell division cycle 7
cDNA	complimentary DNA
DBF4	dumbbell forming factor 4 protein
DBF4	dumbbell forming factor 4 gene
$DBF4_{as}$	antisense ncRNAs at the *DBF4* promoter region
$DBF4_{ups}$	ncRNAs upstream of the DBF4 transcription start site
DHFR	dihydrofolate reductase
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tags
lncRNA	long non-coding RNA
OBR	origin of bidirectional replication
ORC	origin recognition complex
MCM2	minichromosome maintenance complex component 2
min	minute(s)
µl	microlitre
mM	millimolar
µM	micromolar
mRNA	messenger RNA
MtRPOL	mitochondrial RNA polymerase
ncRNA	non-coding RNA
PBS	phosphate buffered saline
poly-A	poly-adenosine
pre-RC	pre-replication complex

RN18S	ribosomal protein 18S
RNA	ribonucleic acid
siRNA	small interfereing RNA
RNAi	RNA interference
RNAP	RNA polymerase
rRNA	ribosomal RNA
TBE	tris-borate-EDTA
tRNA	transfer RNA
SLC25a40	solute carrier family 25, member 40 gene
TSS	transcription start site
UBC	ubiquitin ligase C
Xist	X-inactive specific transcript

1.0 Introduction

1.1 Background

DNA replication is the keystone process of the cell that allows for life to replicate itself and propagate indefinitely. Any dividing cell must go through a specific sequential set of steps which constitutes the cell cycle. The eukaryotic cell cycle is divided into four phases: the G1, S, G2 and M phases (Schafer 1998). Most cycling eukaryotic cells spend the majority of the cell cycle in the first gap phase, G1, which is a period of growth and general preparation for progress into S phase. The beginning of S phase, or synthesis phase, is marked by the initiation of DNA replication and ends once the cell has produced the entire copy of its own genome. The G2 phase, a shorter gap phase, then follows and the cell continues to produce organelles and other cellular constituents until it is ready to divide. Finally, the M phase is the shortest and most complex phase where mitosis occurs; the duplicated DNA is segregated and cells divide in a way that assures both daughter cells will have a full copy of the original genome (Schafer 1998). Quiescent cells remain either temporarily or permanently in the G0 phase and behave like an extended G1 phase, except they typically do not grow in size or produce extra organelles in preparation for the next cell division (Blagosklonny 2006).

1.1.1 The Mechanism of Cell Cycle Control

The eukaryotic cell cycle is largely governed by the cyclins. Cyclins are cyclically expressed proteins that are degraded in a strictly regulated sequence which assures that cells progress in a single direction through the cell cycle in a clockwork manner (Johnson & Walker 1999). Cyclins form complexes with cyclin-dependent kinases (cdks). Cdks are serine-threonine

kinases that, when bound to different cyclins, phosphorylate various key proteins responsible for activating various cell cycle-related tasks.

A large portion of the cell machinery is devoted to assuring that DNA replication initiates in a timely and accurate manner to ensure faithful replication only once per cell cycle (Arias & Walter 2007). Cell cycle regulation is accomplished through the form of checkpoints working in conjunction with the cyclins. Different checkpoints monitor different parts of the cell cycle and, when activated, they prevent further progression into the cell cycle until the source that triggered the checkpoint has been addressed (Elledge 1996; Johnson & Walker 1999). If an actively cycling cell is held for an extended period of time outside of the G1 phase, indicating an inability to rectify the problem in a timely manner, the transcription factor p53 triggers downstream targets eventually leading to the release of active caspases into the cytoplasm resulting in cell death via apoptosis (Elmore 2007).

Defects among the multiple safeguards that prevent abnormal replication are crucial for the transition from normal to cancerous cells, making the mechanics of DNA replication an attractive target of cancer research (Kastan & Bartek 2004). A large portion of this regulation ultimately influences the proteins associated with replication origins – sites at which DNA replication and cell cycle regulating proteins gather and assemble and begin the process of copying DNA when the cell is ready for cell division (Méchali 2010).

1.1.2 Replication origins

The replication origins of prokaryotes, and simple eukaryotes such as budding yeast, are defined by specific consensus sequences signaling precisely where replication origins are located within their genome (Méchali 2010). Mammalian replication origins are also organized at

specific sites throughout genomic DNA, however these sites are more flexible and loosely defined. The lack of consensus sequences for replication origins in most eukaryotes has made it difficult to locate and study them as there is still no method to reliably predict where one of these replication origins is located (Gilbert 2001).

Until recently, only about 30 mammalian origins had been found (reviewed by Aladjem et al. 2006). In the past few years however, recent advances in genome wide replication mapping have uncovered the loci of a large quantity (nearly 14,000) of replication origins in humans (Dellino et al. 2013). It is estimated that to copy the 3,300 megabases of human genomic DNA during the typical 5-6 hour S phase window, 30,000-50,000 active replication origins must fire during every cell cycle, indicating that there are still many origins to be found and characterized (Méchali 2010). Furthermore, only a couple of the mammalian origins, specifically the origins found at the *DBF4* and *Lamin B2* loci, have been described in-depth at the single nucleotide level as there are only a few methods available to study mammalian origins in detail. The first method is the more tedious ligation-mediated PCR method used to characterize the *Lamin B2* origin (Abdurashidova et al. 2000) and the other is replication initiation point (RIP) mapping technique developed in our lab to study the origin at *DBF4* (Romero & Lee 2008b). Currently, neither method has been adapted for high throughput analysis of replication origins, preventing large scale studies of known replication origins at high resolution.

The *DBF4* gene, located on human chromosome 7, codes for the DBF4 protein which is a key regulator of cell-cycle progression. DBF4, or dumbbell forming factor 4, is named for the characteristic dumbbell shape observed in cells when this protein undergoes a loss of function mutation in yeast and is also known as Activator of S phase Kinase (ASK). There is also a *DBF4* pseudogene located on chromosome 10. A pseudogene is a non-functional copy of a gene found

elsewhere on the genome likely due to an abnormal chromosomal duplication event resulting in the generation of non-functional gene segments. Alternatively, pseudogenes may also arise from a retrotransposition, where RNA strands are randomly reverse transcribed into DNA by a reverse transcriptase and, by chance, subsequently inserted into the genome. As the *DBF4* pseudogene lacks the intronic sequences and promoter sequence from the original gene, it would most likely have been created by a retrotransposition event. The *DBF4* pseudogene is also devoid of a replication origin such as the one found at the *DBF4* locus. The pseudogene sequence also lacks the entire 11th exon of the original *DBF4* gene. Apart from the missing exon, however, the rest of the *DBF4* pseudogene is nearly identical in sequence to the exonic region of *DBF4* resulting in important considerations when choosing PCR primers to study the *DBF4* region.

DBF4 protein levels increase and stabilize at the G1/S transition and remain high until the end of S phase (Kumagai et al. 1999). When DBF4 protein levels are high, it forms a complex with Cdc7 which activates Cdc7 kinase functions in a manner analogous to cyclin-Cdk complexes. The active Cdc7-DBF4 complex tethers itself to the pre-RC complex and phosphorylates MCM proteins, activating the helicase function of the MCM complex, and is the final step to signaling cells to begin copying DNA in S phase (Jiang et al. 1999). Cdc7 protein levels remain constant throughout the cell cycle; however, its activity is dependent on DBF4 protein levels (Kumagai et al. 1999).

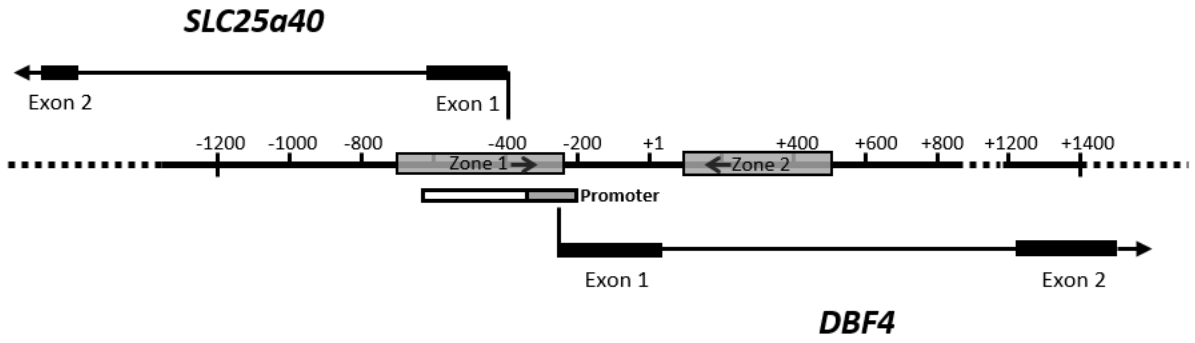
The Origin Recognition Complex (ORC) is a 6-subunit protein complex that can bind to many regions of DNA, marking them as potential replication origins (DePamphilis 2003). In late M/early G1, chromatin-bound ORC recruits Cdc6 to an origin which subsequently recruits Cdt1 and the MCM helicase complex (Bell & Dutta 2002). Together, these proteins form the pre-replication complex (pre-RC). Multiple factors appear to be involved in determining the regions

that ORCs will target. At the beginning of S phase, Cdc6 is phosphorylated and degraded and Cdt1 is inhibited by Geminin (Ballabeni et al. 2004). When cells are ready to progress into the S-phase, Cdc7-DBF4 phosphorylates multiple subunits of the MCM complex leading to the activation of its helicase activity (Weinreich & Stillman 1999; Jiang et al. 1999). The rest of the replisome is then recruited and DNA replication begins.

Only a portion of ORC-bound origins fire in a given cell cycle (Rhind 2006). Some origins fire consistently across all cell types within a species while firing of others varies greatly. Furthermore, after the onset of S-phase, certain origins fire almost immediately (early-firing origins) while others don't begin replicating DNA until later during S phase (late-firing origins) (Méchali 2010). It has been thought that different configurations in heterochromatin as well as the position of DNA within the nucleus across various cell types contributes to the variation in location and timing of active origins (Vogelauer et al. 2002), although the exact regulatory mechanism responsible is not well understood.

Our lab has previously shown that the *DBF4* locus contains an early-firing replication origin in its promoter region and with two ORC-binding sites. What is particular about this origin, is that it contains two initiation zones that fire sequentially in opposite directions in a precisely coordinated manner (Romero & Lee 2008a) (**Figure 1**), a process termed Asymmetrical Bidirectional Replication (ABR). This process is different from the previously studied Origin of Bidirectional Replication (OBR) model, characterized at the human *lamin B2* locus, in which replication was observed to initiate from a single point and extend in both directions at once (Burhans et al. 1990). However, when using the RIP-mapping technique on the *lamin B2* origin, our lab found that the *lamin B2* origin also has a similar gap between replication initiation points in each direction, suggesting that the *lamin B2* origin follows the ABR

Figure 1. Map of the ABR origin at the *DBF4* locus



A map of the *DBF4* promoter region located on chromosome 7. The first two exons of the *DBF4* and *SLC25a40* genes are depicted. The promoter of *DBF4* is located within the indicated boxed region and the core promoter is colored in grey. In the ABR model, DNA replication begins first from multiple sites located within initiation zone 1 into the same direction of *DBF4* transcription and afterwards DNA replication begins on the other DNA strand within initiation zone 2 traveling in the opposite direction. Each zone contains its own set of ORCs and replication proteins. The ladder indicates base positions relative to the ATG (A is +1) sequence of the *DBF4* coding region.

model (Lee & Romero 2012). Another study on the II/9A origin of *Sciara coprophillia* also found a similar separation between replication initiation points (Bielinsky et al. 2001) adding further support to the prevalence of the ABR model in eukaryotic cells.

1.2.1 Long non-coding RNA (lncRNA)

The most well-known role of RNA is its role as a messenger, carrying information transcribed from DNA that can be read and translated into proteins by the ribosomes. However, the majority of RNA is non-coding: this includes the 20 or so transfer RNA (tRNA), the 3 major ribosomal RNAs (rRNA), a variety of ribonucleoproteins and a large number of mRNA of various sizes which do not code for proteins. tRNAs and rRNAs were originally thought to be exceptions to RNA's role as a messenger. RNA whose sequences were found to lack proper open reading frames that can code for protein were once disregarded as corresponding to non-functional by-products of gene transcription that would be eventually degraded by RNAses and recycled for more transcription (Mercer et al. 2009). More recently, however, it has been shown that a large portion of these non-coding transcripts are in fact functional and often associate with proteins allowing them to target various processes in the cell (Mercer et al. 2009). It is currently unclear if most non-coding RNAs (ncRNAs) are functional. However, the list of ncRNAs associated with functions is growing rapidly as effective techniques for determining targets of ncRNAs are now available (Mercer et al. 2009).

ncRNAs are classified based on their size (e.g. long non-coding RNA, microRNA), cellular location (small nucleolar RNA, piwi associate RNA), function (small interfering RNA), shape (short hairpin RNAs) or even location on DNA (long intergenic ncRNA, antisense lncRNA) (Ma et al. 2013). Among these classifications, long non-coding RNA (lncRNA) is a

relatively vague class of RNAs as it encompasses all ncRNAs over 200 bases in size. This minimum length is somewhat arbitrary which is based on the differences in techniques in isolating different size ranges and there is much overlap in function between ncRNAs and lncRNAs.

RNA is capable of forming DNA-RNA triplexes with double-stranded genomic DNA as well as DNA-RNA hybrid molecules or RNA-RNA duplexes forming secondary and tertiary structures that allow them to take on various functional shapes and be recognized by proteins (Mercer et al. 2009). This provides an interface for proteins bound to ncRNAs and allows them to target DNA or RNA sequences within the cell with high sequence specificity. The same protein is able to recognize different targets depending on the ncRNA it is bound to which permits more variety in targets that bind to a single protein. Many lncRNAs have been implicated recruiting chromatin modifying enzymes to DNA (Mercer et al. 2009). The well-known *Xist* lncRNA coats the inactive X chromosome and recruits chromatin-modifying enzymes that recognize *Xist* and silence the X chromosome from which *Xist* is transcribed (Chow et al. 2005). Other lncRNAs are also able to recruit chromatin modifying enzymes throughout the genome acting as epigenetic regulators (Lee 2012).

LncRNAs regulate transcription activators and repressors, as well as RNA polymerase II (RNAP II), providing yet another level of transcriptional regulation (Goodrich & Kugel 2006). Additionally, genes are regulated by lncRNAs with hairpin structures which can be cleaved into smaller fragments by DICER and subsequently used as small interfering RNA (siRNA). SiRNAs bound to the RNA-induced silencing complex (RISC) act as a template to target other complimentary transcripts to be degraded thereby reducing the transcript pool for individual proteins (Tijsterman & Plasterk 2004). Furthermore, the act of transcription of lncRNAs can

simply affect nearby genes by opening up chromatin to allow easier access for transcription factors to bind to regulatory regions of other genes or to compete for the same regulatory region (Magdalou et al. 2014).

1.2.2 RNA transcription

There are 4 known RNA polymerases found in the mammalian system. Each of the mammalian RNA polymerazes recognize different sets of DNA sequences with the aid of various transcription factors allowing them to transcribe specific RNA products. RNA polymerase I (RNAP I) is responsible for transcribing the sequences for the 5.8S, 18S and 28S ribosomal RNAs (rRNAs) (Drygin et al. 2010). RNAP I is located uniquely to the nucleolus where it transcribes the rRNAs which are subsequently cleaved during post-translational modifications to generate the three individual rRNA molecules. The human genome contains hundreds of copies of the rRNA gene transcribed by RNAP I. Despite the low number of RNAP I targets, they are the most actively transcribed genes and constitute approximately 80% of total cellular RNA (Drygin et al. 2010).

RNAP II is responsible for the transcription of the protein-encoding mRNAs in the nucleus as well as most ncRNAs (and lncRNAs) (Hahn 2004). This polymerase is distinguished by its C-terminal tail on its largest subunit, where many proteins necessary for processing RNAP II-transcribed RNAs reside (Phatnani & Greenleaf 2006). Among these proteins are methyltransferases, which are responsible for capping RNAP II products with guanine at the 5' end. The 5' guanine cap is essential in providing degradation resistance to mRNAs from cytosolic proteins that degrade RNA and is also important for mRNA transport and translation (i.e., RNases) (Cho et al. 1997). Another feature of RNAP II products not found with RNAP I or

III is that they have a poly-A tail which consists of multiple repeats of adenine at the 3' end of its RNA products (Eggermont & Proudfoot 1993). These poly-A tails add additional stability to mRNAs. As poly-A tails are degraded over time in the cytosol, the coding regions on the Mrna are eventually exposed and are degraded resulting in a non-functional transcript (Slomovic et al. 2010). Another notable protein complex that associates with the RNAP II C-terminal tail is the spliceosome, which is responsible for removing introns from RNAP II-transcribed mRNA (Will & Lührmann 2011).

Similar to RNAP I, RNAP III is involved in the transcription of a narrow spectrum of RNA species. RNAP III is responsible for the transcription of all tRNAs as well as the 5S rRNA, which is the only rRNA not transcribed by RNAP I (White 2011). This polymerase is also responsible for transcribing small interspersed nuclear elements (SINEs), one of the most abundant elements in the human genome, among other ncRNA and lncRNA (Cordaux & Batzer 2009).

Lastly, the mitochondrial RNA polymerase (MtRPOL), which is transcribed in the nucleus by RNAP II, is thought to be exclusively located to the mitochondria where it transcribes all of the mitochondrial DNA (Arnold et al. 2012). However, some recent studies indicated that MtRPOL, or perhaps a truncated version named spRNAP-IV, may localize to the nucleus. This nuclear form of MtRPOL is reported to transcribe some genes in the nucleus including *ALDH8* (Kravchenko et al. 2005) as well as a number of muscle actin genes (Lee et al. 2011). There has been very little follow-up on the potential nuclear activities of MtRPOL and it is not known how a nuclear MtRPOL would be able to recognize transcription sites on chromatin which, unlike mitochondrial RNA, is more tightly packed with histones and other proteins. Most recently,

however, it has been suggested that MtRPOL does not locate in the nucleus or transcribe *ALDH8* (Kühl et al. 2014).

Many groups have noted a possible interplay between transcription and replication cellular machinery, as both of these processes must physically cross paths on DNA when the duplication of the cell's genetic content is under way (Schneider & Grosschedl 2007; Giri & Prasanth 2012). Interestingly, our lab has previously noted that *DBF4* transcription levels temporarily decrease during high replication origin activity and increase after replication in the area is complete (Romero 2008). During the process of characterizing the *DBF4* replication origin we surveyed this region for the presence of non-protein coding transcripts. In this thesis I report the presence of multiple lncRNAs located within the two ORC binding sites at the *DBF4* locus. The objectives in this research are to confirm that these lncRNAs are transcriptionally different than the *DBF4* gene with which it overlaps, to determine which RNA polymerase is responsible for transcribing these lncRNAs and to determine if transcription levels of the lncRNAs have an effect on neighbouring genes on the human chromosome.

2.0 Materials and Methods

2.1 Cell culture

HeLa S3 (ATCC CCL-2.2; Manassas, Virginia, USA) cells were cultured in RPMI-1640 medium containing 10% v/v fetal bovine serum (Hyclone – GE Life Sciences; Mississauga, Ontario, Canada). Culture of HEK293T (ATCC CRL-3216) cells was carried out using Dulbecco's modified Eagle's medium containing 10% v/v fetal bovine serum. Human mammary epithelial 184B5 (ATCC CRL-8799) cells were cultured in serum-free, mammary epithelial basal medium (MEBM) supplemented with 1 ng/ml cholera toxin, 0.5 μg/ml hydrocortisone, 10 ng/ml human epidermal growth factor, 5 μg/ml insulin, 13 μg/ml bovine pituitary extract and 1% v/v GA-1000 (gentamicin-amphotericin mix, Lonza, distributed through CEDARLANE; Burlington, Ontario, Canada). All cells were incubated in a humidified environment at 37°C with 5% CO_2. When needed, α-amanitin was aseptically mixed into the media at a concentration of 111 μg/ml.

2.2 Synchronization

HeLa cells were synchronized at the G1/S border using a double-thymidine block. Briefly, asynchronous HeLa cells were incubated in culture medium containing 2 mM thymidine (Sigma) for 18 h. Cells were then washed once with 1x PBS pH 7.4 (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) and released into culture medium for 11 h. Cells were then incubated a second time in regular medium containing 2 mM thymidine for 13.5 h. After another wash with PBS pH 7.4, cells were released in culture medium and samples were collected at the indicated times for analysis.

2.3 Flow cytometry

Cell cycle progression was monitored using flow cytometry to measure the DNA content of cells. Cells were harvested by trypsinization, washed once with PBS pH 7.4, and fixed by incubation in 70% v/v ethanol for at least 18 h at -20°C. The cells were then washed twice with PBS pH 7.4 and incubated in propidium iodide staining solution (0.1% w/v sodium citrate, 0.3% v/v NP-40, 100 µg/ml propidium iodide, 100 µg/ml RNase A) for >1 h at room temperature. Samples were analyzed using a Beckman Coulter FC500 flow cytometer.

2.4 RNA isolation and analysis of RNA using primer extension

RNA was isolated using a QIAGEN RNeasy Mini kit following the manufacturer's protocol (QIAGEN, Mississauga, Ontario, Canada). RNA concentrations and purity were determined by measuring the OD260 and OD280 using the NanoDrop 2000 (Thermo Scientific; Ottawa, Ontario, Canada). 5 µg of RNA, 6 µl of 5x first strand buffer (Invitrogen; Burlington, Ontario, Canada) and 5 pmol of DIG-labeled primer (Roche; Mississauga, Ontario, Canada. see **Table 1**) were mixed together and RNase-free water was added to a final volume of 23 µl and vortexed very lightly. The mixture was then brought to the 65°C annealing temperature for 10 min and then cooled on ice. 3 µl of 8 mM dNTPs, 3 µl of 0.1 M DTT and 1 µl of Superscript II reverse transcriptase (Invitrogen; Burlington, Ontario, Canada) were added to the mixture for a final volume of 30 µl. Samples were then incubated for 1 h at 42°C for extension. Once this procedure was completed, the reverse transcriptase was denatured by incubating at 90°C for 5 min, and chilled on ice, followed by the addition of 20 µl formamide gel loading buffer (98% formamide; 10% EDTA pH 8.0; 0.1% xylene cyanol; 0.1% bromophenol blue). Samples were

heated to 65°C before being added to a 5% w/v acrylamide gel containing 7.7 M urea in 0.5% TBE (Tris-borate-EDTA), and the DNA was then separated by PAGE. The DIG-labeled DNA molecular weight marker V (Roche; Mississauga, Ontario, Canada) was used alongside samples as a size marker. The labeled cDNA was then transferred to a Hybond N+ nylon membrane (GE Life Sciences; Mississauga, Ontario, Canada) using a semi-dry transfer apparatus. cDNA was then fixed to the membrane by exposure to UV light for 6 min. Finally, the extended products were detected using a DIG detection kit from Roche (Mississauga, Ontario, Canada).

2.5 Analysis of transcript using RT-qPCR

qPCR was carried out using a 7900HT ABI Prism thermocycler system (Applied Biosystems; Burlington, Ontario, Canada). Reactions were performed in triplicate in a final volume of 25 µl containing 12.5 µl of 2× SYBR Green Master Mix (Applied Biosystems; Burlington, Ontario, Canada) and 300 nM of each primer (Integrated DNA Technologies; Coralville, IA, USA). Primer sequences are shown in **Table 1**. Standard curves were generated for each primer pair using serial dilutions of total genomic DNA (10 ng, 1 ng, 0.1 ng and 0.01 ng). An equal volume of gel-extracted nascent DNA (usually 2 or 3 µl) was used as template for amplification with every primer set. Following initial denaturation for 10 min at 95°C, amplification was carried out for 40 cycles as follows: 95°C for 30 s, 59°C for 30 s, and then 72°C for 30 s. Amplification of a single product of the correct size was verified by melting curve analysis and by agarose gel electrophoresis. Relative sequence abundance in a given nascent DNA preparation was estimated by calculating the ratio between the amount of each DNA segment (as estimated by standard curves) and the amount amplified by the Prom3 primer set in

the same DNA sample. Primer pair Prom3 was chosen as a control because it consistently showed the lowest levels of amplification.

2.6 Determination of transcript half-life

The half-life of the $DBF4_{ups}$ (ups for upstream promoter sequence) transcript was studied using a Click-iT Nascent RNA Capture Kit (Invitrogen; Burlington, Ontario, Canada). The experiment was preformed twice according to the manufacturer's protocol. Briefly, HeLa cells were incubated in regular medium containing ethynyl uridine (EU) for 24h. 0 h samples of total RNA were collected immediately after release into fresh medium in the absence of EU (EU-) and used as the pre-degradation baseline. A second round of total RNA was collected 24 h after release into fresh medium. Total RNA samples were mixed with biotin provided in the kit which binds to the labeled RNA. Dynabeads® MyOne™ Streptavidin from the kit were then mixed with the total RNA samples and the samples were washed according to manufacturer's protocol to separated labeled RNA from total RNA. Labeled RNA levels between the two samples were then analyzed using the 7900HT ABI Prism thermocycler system. As the first attempt generated signal near the non-linear edge of the standard curve generated during qPCR, the experiment was repeated in triplicate using a 12 h time period instead of 24 h.

2.7 Knockdown of MtRPOL using siRNA

Cells were transfected with siRNA using the Lipofectamine 2000 transfection reagent (Invitrogen; Burlington, Ontario, Canada) in OptiMEM medium (Life Technologies; Burlington, Ontario, Canada). The experiments were preformed according to the manufacturer's protocol

(Invitrogen; Burlington, Ontario, Canada) scaled to 10 cm cell-culture plates. For transfection experiments, cultured HEK293T cells were used since the transfection efficiency was generally over 90%. SilencerSelect siRNA directed towards MtRPOL (Life Technologies, cat# AM16708) was used in this experiment in conjunction with SilencerSelect negative control No. 1. Western blot was done to confirm knockdown of MtRPOL using an antibody directed towards MtRPOL (cat# SC-365082, Santa Cruz; Dallas, Texas, USA) with antibodies for DBF (cat# ab124707, Abcam; Cambridge, Massachussets, USA) 4 and CRM1 (cat# SC-5595, Santa Cruz) used as loading controls.

3.0 Results

3.1.1 Multiple lncRNA are found surrounding the *DBF4* promoter region

To search for the transcription start sites (TSSs) of transcripts near the promoter region of *DBF4*, primer extension was carried out on total RNA isolated from HeLa cells with digoxigenin (DIG) labeled primers. DIG is a steroid isolated from *Digitalis* sp which can be very specifically targeted with an anti-DIG antibody. The antibody is then linked to the chemiluminescent substrate CDP-Star, permitting highly sensitive detection of labeled primers without requiring PCR to amplify them to obtain detectable bands on a gel. The labeled primers targeted the 5' ends of putative transcripts that can be transcribed in the *DBF4* promoter region and were extended into cDNA using reverse transcription. The cDNA is a reverse-complement copy of the RNA synthesized by reverse transcriptase (RT) starting from where the primer targets (listed in **Table 1**) until the RT enzyme reaches the 5' end of targeted RNA strand. cDNA synthesis is terminated when the RT reaches the end of an RNA strand, producing products of a specific length. The cDNA was then resolved by gel electrophoresis, followed by blotting and detection (**Figure 2A**). As there is no PCR amplification, the resulting bands represent directional transcripts, preventing accidental detection from antisense targets such as the neighboring *SLC25a40* gene. As the sequence of the original primer is mapped to the human genome, the rest of the extended 5' sequence can be deduced based on the size of the products, assuming the template RNA is not spliced.

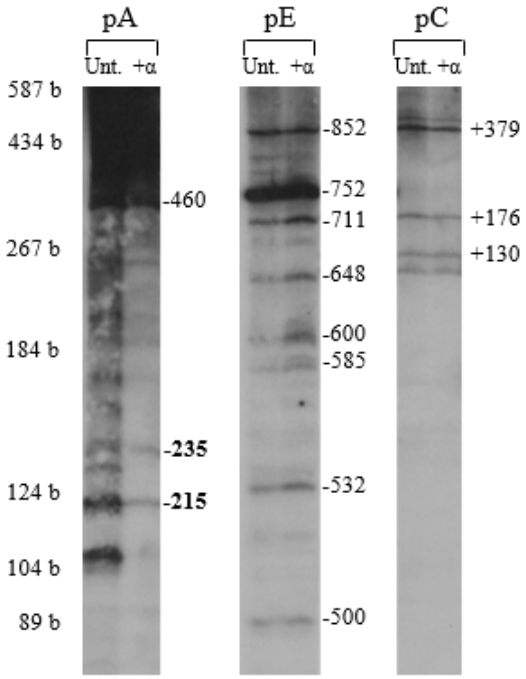
As can be seen in **Figure 2B**, Primers pA and pE target the 5' ends of the RNA transcribed in the same direction as *DBF4* (primer sequences are found in **Table 1**). Primer pA starts at position -106 (i.e., upstream) relative to the ATG of *DBF4* and is thus able to detect the

Figure 2. Multiple lncRNAs were detected at the *DBF4* locus

A. Primer extension was used to detect lncRNA in the *DBF4* promoter locus. The position of the most prominent bands relative to the ATG start codon of the *DBF4* gene have been labeled to the right while the molecular size markers are listed on the far left. In bold under the pA primer are the major transcription start sites of the *DBF4* mRNA and as expected their expression levels were greatly reduced in the presence of RNA polymerase inhibitor. pA and PE target transcripts that are sense to the *DBF4* gene while pC targets transcripts antisense to *DBF4*. A high exposure is shown to visualize the lower-sized bands.

B. The map constructed with data obtained from Panel A indicating the 5' ends of the lncRNAs as determined using primer extension. The transcription start sites shown are mapped assuming that the detected portion of the transcripts were not spliced. The start positions of exon 1 of *DBF4* and *SLC25a40* and the end position of *SLC25a40*'s exon are indicated. Two PCR primers were used for $DBF4_{ups}$ detection: Prom C7UC which amplifies the -640 to -703 region and Prom 7B which targets -625 to -675. These primers were chosen because neither of them locate within *SLC25a40* exons. DIG-labeled primers for primer extension as well as PCR primer sets that target the lncRNAs are also shown. All distance values are relative to the ATG (+1) of the *DBF4* translational start site.

A


pA
pE
pC
Unt. +α
Unt. +α
Unt. +α
587 b
434 b
267 b
184 b
124 b
104 b
89 b
-460
-235
-215
-852
-752
-711
-648
-600
-585
-532
-500
+379
+176
+130

B

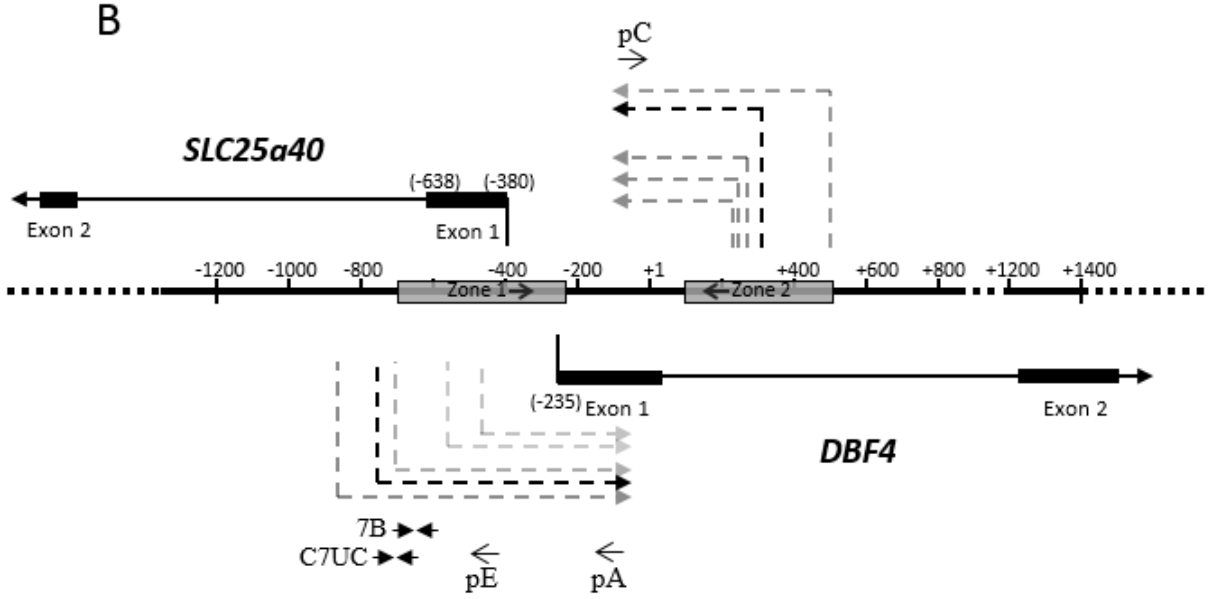

pC
SLC25a40
(-638)
(-380)
Exon 2
Exon 1
-1200
-1000
-800
-400
-200
+1
+400
+600
+800
+1200
+1400
Zone 1
Zone 2
(-235)
Exon 1
Exon 2
DBF4
7B
C7UC
pE
pA

Table 1. Digoxigenin-labeled primers used to detect the 5' end of lncRNAs

Oligo	Sequence (5'-3')	Size (bases)	% GC	Position to ATG	Tm (°C)
pA	CTTCCGCCAGCTACGGCCTC (S)	20	70.0	-106 to -125	68
pC	GCGTAGAGGCCGTAGCTGGC (AS)	20	70.0	-130 to -111	68
pE	AGGTTGTGTTTCCGCCTCTAC (S)	21	52.4	-402 to -422	64
GAPDH	CCATGTAGTTGAGGTCAATG	20	45.0	n/a	58
chr10a	GAGTGAACATTCTGAGATTTAT (AS)	22	31.8	-556 to -535	50
chr10b	GGAGTCTTATATGAACTTATG (S)	21	33.3	-609 to -629	46
chr10c	GGCAGAAAATCTAAACCATTGA (S)	22	36.4	-334 to -355	58

Primers are labeled with digoxigenin at the 5' end. The indicated positions of primers pA, pC and pE relative to the ATG sequence of *DBF4*. The indicated positions of ch10a, ch10b and ch10b are relative to the ATG sequence of the *DBF4* pseudogene. Primers are indicated whether they target sense (S) or antisense (AS) direction to *DBF4* or the *DBF4* pseudogene

DBF4 transcription start sites (TSSs) located -215 and -235 from the ATG. Primer pE targets sense transcripts further distally from the *DBF4* TSS at -402 allowing for better resolution of the $DBF4_{ups}$ transcripts. Primer pC is complimentary to the 5' ends that would be transcribed antisense to *DBF4*. The resulting blot reveals the presence of the TSSs for the *DBF4* gene (-215 and -235) (**Figure 2A**). Furthermore, several bands were observed indicating the presence of TSSs upstream to the *DBF4* gene transcript between -500 and -850 bases from the *DBF4* ATG sequence. These *DBF4* upstream transcripts ($DBF4_{ups}$) were not sensitive to treatment with the RNAP II inhibitor α-amanitin (**Figure 2A** and **Figure 3**). LncRNAs were also observed between +100 and +400 in the antisense direction to the *DBF4* gene ($DBF4_{as}$) with the use of primer pC, which were resistant to α-amanitin. Several other transcripts even further from the ATG sequence in both directions were detected, however these larger-sized transcripts displayed sensitivity to α-amanitin, suggesting that these sequences are not of the same nature as the $DBF4_{ups}$ transcripts and the $DBF4_{as}$ transcripts. **Figure 2B** summarizes the directions, relative strengths and putative locations of the detected α-amanitin-insensitive lncRNA transcripts.

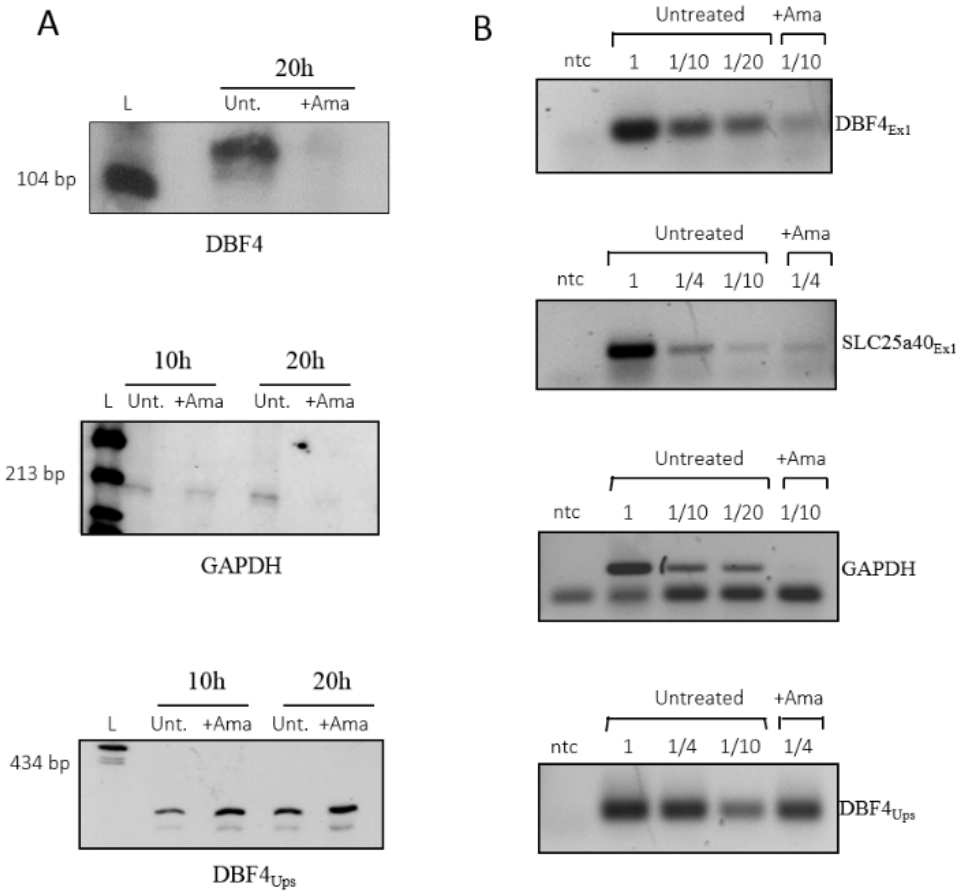
3.2.1 LncRNAs in the *DBF4* promoter region are differentially transcribed from neighboring genes

A common feature of transcribed RNA is the presence of splice sites – regions of RNA that are cut out from newly-synthesized mRNA sequences during, or shortly after, the transcription process. Splicing results in discontinuous regions of DNA being represented in the final transcripts (e.g., mRNA). While primer extension experiments yield bands corresponding to the distance from TSSs for the targeted RNA segments, it does not indicate if these bands correspond to a continuously-transcribed sequence of the *DBF4* region or if splice sites are found within the RNA segment as the extended RNA sequence is still unknown. Furthermore, a single

gene such as *DBF4* may have alternate splice sites that are not always spliced. Alternative splice sites result in multiple variants of RNA species from a defined region of the genome, often producing non-protein-coding RNA products alongside its coding counterpart (Luco & Misteli 2011).

To differentiate whether the *DBF4*$_{ups}$ transcripts are indeed novel lncRNAs or if they are simply alternate transcripts of *DBF4,* HeLa cells were treated with the RNA polymerase II (Pol II) inhibitor α-amanitin for 20 h. *DBF4* is transcribed by Pol II and its expression levels should therefore be sensitive to Pol II inhibitors. Using DIG-labeled primer extension, I observed that the *DBF4*$_{ups}$ lncRNAs (**Figure 3A**) and the *DBF4*$_{as}$ lncRNAs (**Figure 3**, primer pC) showed comparable, and possibly increased, signal levels when compared to the non-treated control. When targeting the *DBF4*$_{ups}$ region (primers targeting -675 to -625 from the *DBF4* TSS) using conventional PCR, the same trend can be seen (**Figure 3B**). The lack of apparent reduction in lncRNA level suggests that the *DBF4*$_{ups}$ and *DBF4*$_{as}$ transcripts are not transcribed by RNAP II. In contrast, the Pol II-transcribed *GAPDH* and *DBF4* genes were reduced in expression levels by 20 h post-treatment. The *SLC25a40* gene, located immediately upstream and antisense to *DBF4*, also showed reduced expression levels in the presence of α-amanitin (**Figure 3B**), indicating that the lncRNAs detected are not products of alternate transcription of *SLC25a40*. **Table 2** shows the list of primers used for all conventional PCR and qPCR experiments.

Figure 3. RNAP II inhibitor α-amanitin does not reduce *DBF4$_{ups}$* expression levels



A. HeLa cells were treated with 111 µg/mL α-amanitin for 10 or 20 h to determine if lncRNA was transcribed by RNAP II. Primer extension was carried out using total RNA from treated (+ama) and non-treated (unt) samples with DIG-labeled primers mapping to the *DBF4* TSS as well as the RNAP II-transcribed *GAPDH* gene. Bands corresponding to the TSSs for GAPDH and *DBF4* exhibited downregulation by 20 h post-treatment. Conversely, the band corresponding to the DBF4$_{ups}$ transcripts (shown is the -752 start site) had stronger signal intensity in the α-amanitin treated sample. Primer sequences are indicated in **Table 1**

B. Total RNA from HeLa cells treated with 111 µg/mL α-amanitin were also analyzed using RT-PCR to confirm the accuracy of the primer extension method. In accordance with primer extension, primers mapping the first exon of *DBF4* and *SLC25a40* (DBF4$_{ex1}$ and SLC25a40$_{ex1}$ respectively) as well as *GAPDH* were all downregulated in the presence of α-amanitin. The primer mapping the DBF4$_{ups}$ region between *SLC25a40* and *DBF4* did not show reduction in expression. Numbers above each sample correspond to dilutions used for semi-quantitative analysis of the 1/10 dilution used to compare the two samples. NTC, no template control. PCR Primer sequences are indicated in **Table 2**.

Table 2. The sequences of all primer sets used for PCR, RT-PCR and qPCR

Target	Primer Name	Sequence (5'-3')	Amplicon Size (bp)	Position (from DBF4 ATG)
$DBF4_{Ups}$	C7UC F	CGGACTCAGCCAAAG	64	-703 to -640 (MCFP intron 1)
	C7UC R	ATGCTAATTTTCCACTGGTC		
$DBF4_{Ex3}$	DBF4 Exon 3F	ACCTTGGGTCGAATTTCTCCTGTACC	88	+796 to +860 (DBF4 exon 3)
	DBF4 Exon 3R	AACTTCCATCATGGCTGGGATGAG		
$MCFP_{Ex1}$	MCFP Ex1 F	TCCATTCCGTTGCTCTCTCAGTCT	175	-429 to -604 (MCFP Exon 1)
	MCFP Ex1 R	TGGCAAACCTAGAAGGCGGGAATA		
RPS28	RPS28-F	AATTCATGGACGACACGAGCC	219	n/a
	RPS28-R	ATCTCAGTTACGTGTGGCGGA		
ND1	ND1F	GAGCAGTAGCCCAAACAATCTC	140	n/a
	ND1 R	GGGTCATGATGGCAGGAGTAAT		
$MtRPOL_{pre-mRNA}$	RNAPOL4F	GTGGTTTCTTATGCAGCCTC	136	n/a
	RNAPOL4R	ATCCTTCTCCAGTATCTTTGC		
Prom3	Prom3 F	ATAAGGACCACGCAACCCAG	95	-31,348 to -31,254
	Prom3 R	GCTCCACCTCCTGTCAGATCA		
Prom7C2	Prom7C2 F	AGTCCTGGCAAACCTAGAAG	80	-608 to -529
	Prom7C2 R	ACCTAGTGGATGCAGTTAGC		
Poly-A cDNA primer	Oligo dT	TTTTTTTTTTTTTTTTTV	n/a	n/a
UBC	UBC F	GCCACGTCAGACGAAGGGCG	70.7	277
	UBC R	CTAGCTGTGCCACACCCGGC	68.4	
RN18S	RN18s F	ATACATGCCGACGGGCGCTG	70.6	80
	RN18s R	AGGGGCTGACCGGGTTGGTT	69.2	

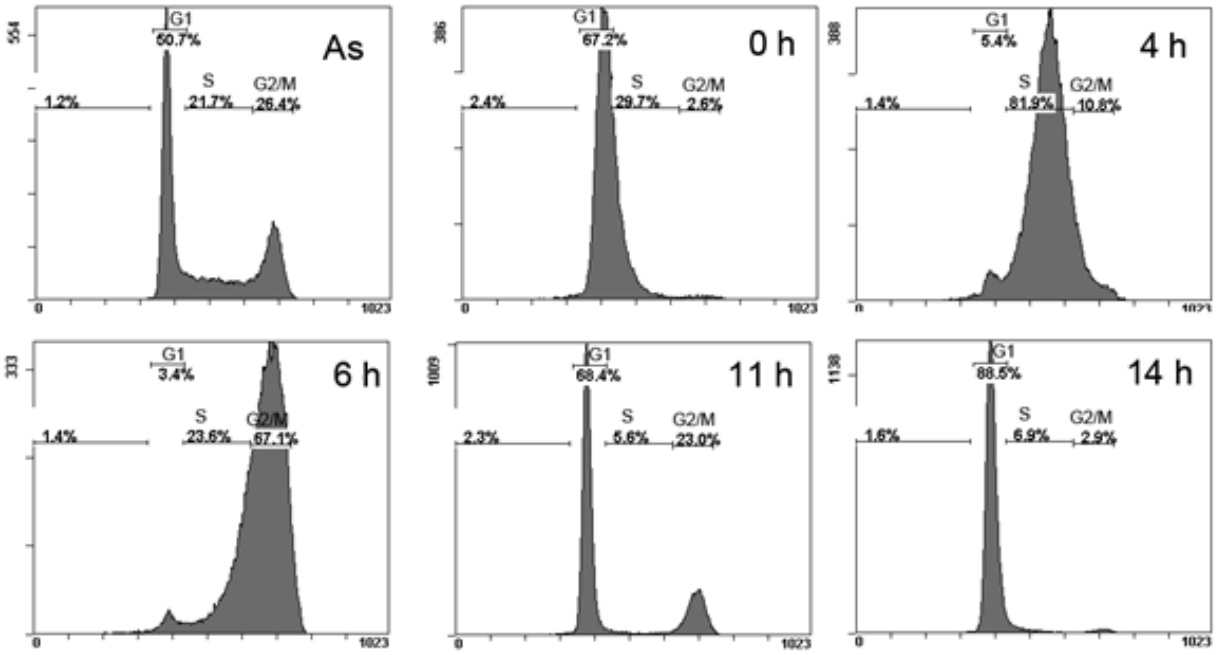
The indicated positions of the primers are positions relative to the ATG sequence of *DBF4* (with "A" as +1). "V" indicates a mixture of primers containing any nucleotide that is not T in that position.

3.2.2 *DBF4$_{ups}$* transcription remains constant throughout the cell cycle

Excess thymidine creates a negative feedback loop preventing cells from producing dCTP. Replication stalls as DNA polymerase no longer has a substrate to incorporate cytosine into nascent DNA causing a cell cycle arrest in S phase when exposed to high concentrations of thymidine. Cells not currently in S phase will continue through the cell cycle and stall as they near the next S phase. Once the excess thymidine is removed from the culture media, the production of dCTP resumes and cells continue through the cell cycle (Bostock et al. 1971). The double thymidine block involves two time periods of excess thymidine to synchronize the majority of cells in a population at early S phase. As shown in **Figure 4**, when HeLa cells are synchronized in the beginning of S phase by double thymidine blocks and then released into complete medium (0 h), they entered the G2/M phase by 6 h, and the G1 phase of the next cell cycle by 11 h (**Figure 4**).

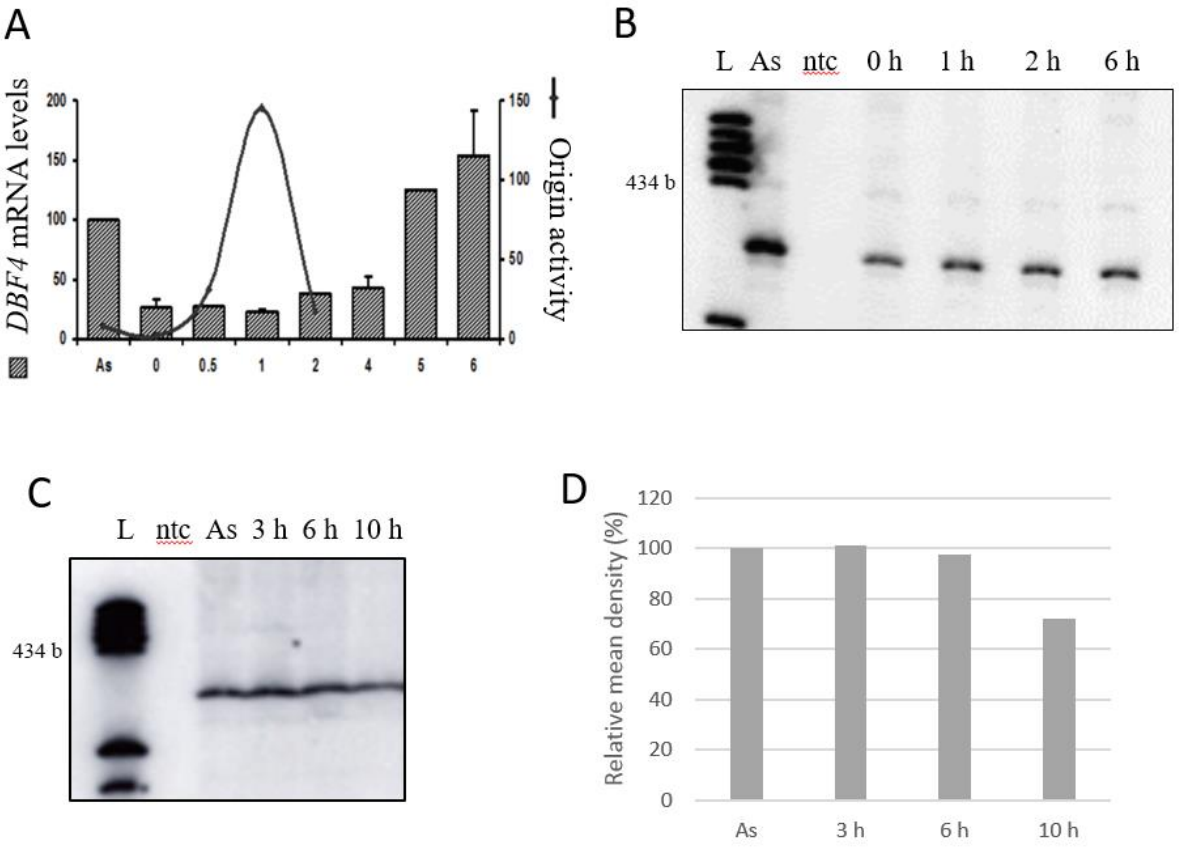
Using the double thymidine block-based cell synchronization, our lab previously noted that *DBF4* expression levels were reduced in early S phase, during the period of high origin activity (Romero 2008). Between 0 h and 4 h after release into the cell cycle, *DBF4* mRNA levels were low compared to the asynchronous population; however, the level increased by 5 h post-double thymidine block and release (**Figure 5A**). HeLa cells were at the G2/M phase by 6-8 h, and reached G1 starting at 9 h post-synchronization. To investigate the expression levels of *DBF4$_{ups}$* as a function of the cell cycle, RNA samples were collected at various cell cycle time points (shown in **Figure 5B** and **5C** are 1, 2, 3, 6 and 10 hour time points) and monitored by qPCR using primer pE. Unlike *DBF4* mRNA (**Figure 5A**), the *DBF4$_{ups}$* expression levels remained constant between the asynchronous population and all time points examined throughout the S

Figure 4. Cell cycle profile of synchronized HeLa cells using double thymidine block



Flow cytometry profile measuring DNA content of synchronized HeLa S3 cells after double thymidine block. A large portion of cells were in the G1 phase in the asynchronous (As) population without block reflecting the larger amount of time spent in the G1 phase by these cells. At 0 h post-double thymidine block, only one peak is seen as cells are still in the beginning of S phase. The peak migrates through the gated S phase and by 6 h the majority of cells in the population contain twice the DNA content of a G1 cell. Cells remain a few hours in G2/M phase and mitosis completes by 11h, where most cells can be seen to have split, and are almost exclusively found in G1 phase by 14 h.

Figure 5. *DBF4$_{ups}$* expression is relatively constant throughout the cell cycle



A. Origin activity, as indicated by the solid line, was measured using the nascent strand abundance assay in the context of DBF4 expression levels. When origin activity is high, expression levels of DBF4 is reduced. Activity levels are measured in % compared to the asynchronous population. (Romero 2008)

B. The levels of *DBF4$_{ups}$* are relatively constant during S phase. Total RNA isolated at 0, 1, 2 and 6 h post-release from HeLa cells synchronized at the G1/S border using double thymidine block were subjected to DIG primer extension using primer pE, followed by acrylamide gel electrophoresis. Shown in the -752 TSS indicated on **Figure 2** for primer pE. L: ladder, As: asynchronous HeLa cells, ntc: no template control.

C. The levels of *DBf4$_{ups}$* were relatively constant throughout S phase up to early G1 of the next cell cycle. The experiment was carried out the same as in panel B, except with different time-points post-G1/S. Primer extension directed to the *DBF4$_{ups}$* transcripts indicated no variation in expression levels at the indicated time points. L: ladder, As: asynchronous HeLa cells, ntc: no template control.

D. Densitometry was used to measure panel C by measuring the relative mean density of each column using ImageJ for each time point. Asynchronous (As) population was normalized to 100%

phase after release from double thymidine-based cell-cycle arrest, however the expression level was reduced slightly in late M phase/early G1 (**Figure 5B and 5C**).

3.2.3 *DBF4$_{ups}$* transcripts are not transcribed by RNAP I, II and III

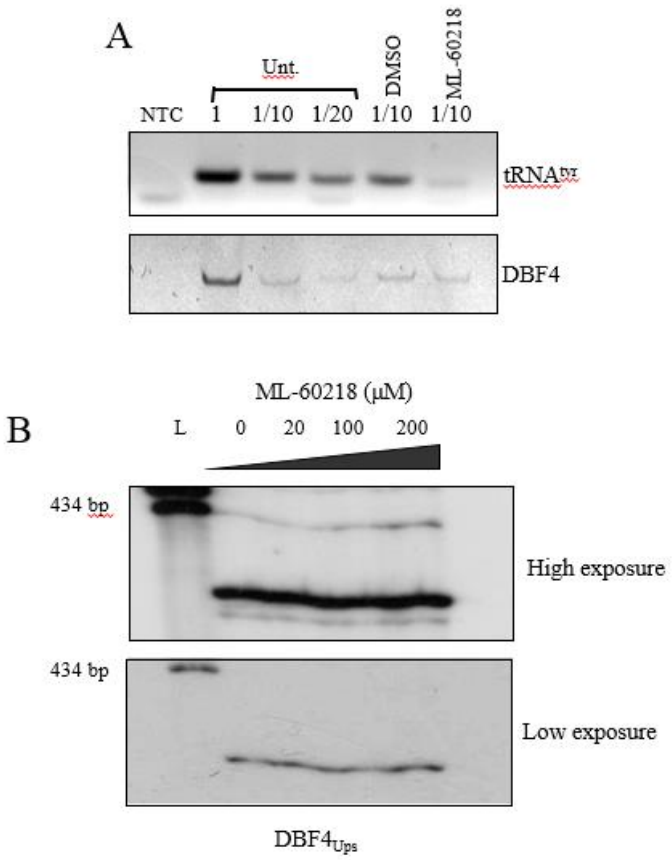
As RNAP II inhibition, using α-amanitin, did not reduce the expression of the *DBF$_{ups}$* transcripts, the possibility that RNAP III was responsible for transcribing the *DBF4*$_{ups}$ transcripts was further investigated as RNAP III is also known to transcribe many lncRNAs. HeLa cells were treated with up to 200 μM of the RNAP III inhibitor ML-60218 over 20 h (**Figure 6**). While the RNAP III-transcribed tyrosine transfer RNA (tRNA^{tyr}), the positive control, was down-regulated in the presence of the RNAP III inhibitor, *DBF4$_{ups}$* expression remained the same.

Although RNAP I is not known for transcribing ncRNAs other than most rRNAs, the effect of RNAP I inhibition was examined mainly because the inhibition of RNAP II and III did not cause any significant down-regulation of the *DBF4*$_{ups}$ transcripts. HeLa cells were treated with 0.04 μg/ml of actinomycin D, which inhibits transcription of RNAP I at lower concentrations and RNAP II at higher doses. As anticipated, the 45S rRNA which is transcribed by RNAP I was downregulated after 3.5 h (**Figure 7**). Similar to the treatment with α-amanitin, however, actinomycin D-treated cells resulted in a stronger *DBF4$_{ups}$* band after 20 h treatment, suggesting that RNAP I, II and III were not involved in transcribing the *DBF4*$_{ups}$ lncRNAs.

3.3 The half-life of the *DBF4$_{ups}$* transcripts is relatively short.

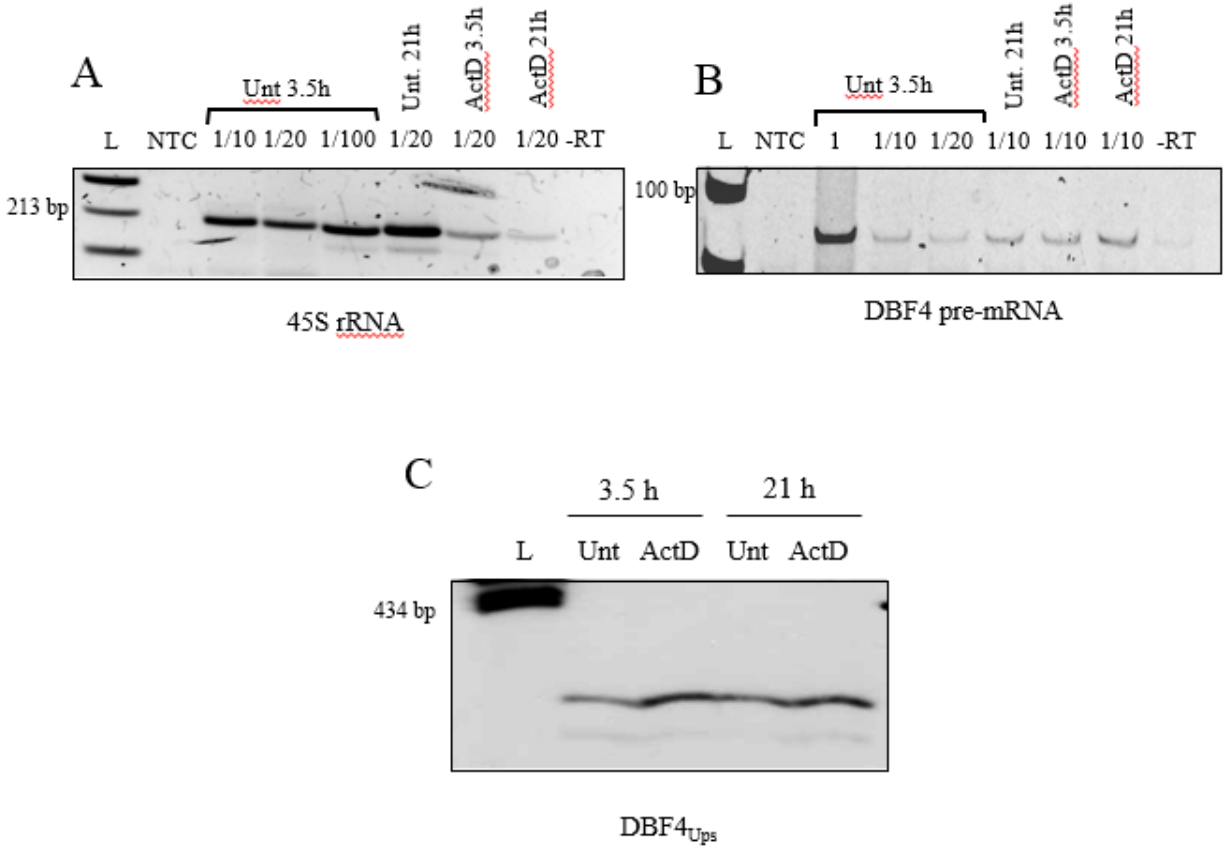
Lack of the down-regulation of *DBF4$_{ups}$* transcription raised the question of whether these transcripts were transcribed by RNAP I, II or III, but down-regulation was not observed in the studied time frame due to these lncRNAs having a particularly long half-life. RNAs isolated

Figure 6. RNAP III inhibitor ML-60218 does not reduce *DBF4$_{ups}$* expression levels



A. HeLa cells were treated with 200 μM of RNAP III inhibitor ML-60218 for 20 h. Serial dilutions of untreated total RNA were included for semi-quantitative analysis of PCR products. When compared the 1/10 dilution of untreated RNA sample with the 1/10 dilution of ML-60218, a reduction in expression can be seen for the RNAP III-transcribed tRNA^{tyr} (top panel) while there is no effect on DBF4 pre-mRNA expression levels (bottom panel). ntc: no template control, DMSO: DMSO-only control
B. Total RNA from cells treated with increasing concentrations of ML-60218 for 20 h were subject to primer extension using primer pE. Short (bottom panel) and long (top panel) exposures of the DBF4$_{ups}$ bands show no reduction in expression level. L: ladder

Figure 7. RNAP I inhibitor actinomycin D does not reduce *DBF4$_{ups}$* expression levels



Total RNA isolated from HeLa cells treated with 0.04 µg/ml RNAP I-inhibitor, Actinomycin D was collected 3.5 and 21 h post-treatment.

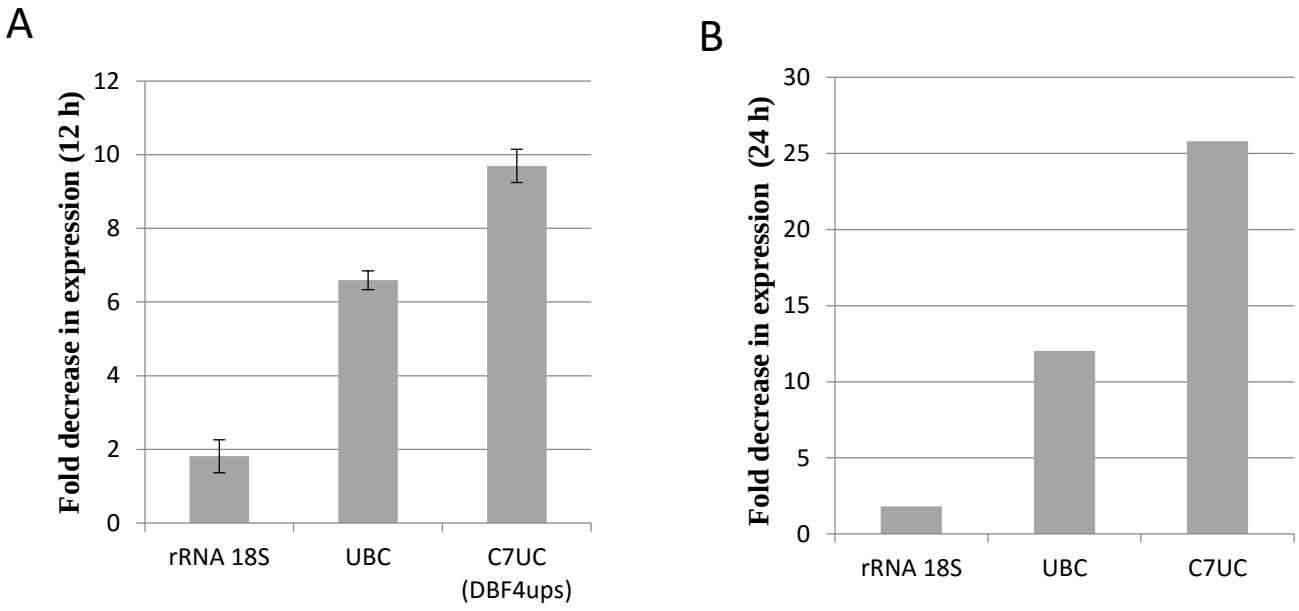
A. PCR for the RNAP I-transcribed 45S rRNA indicated a reduction of expression level when compared to the 1/20 dilutions across samples. ntc: no template control, L: ladder, -RT: samples lacking reverse transcriptase ruling out contaminating DNA
B. cDNA PCR-based amplification of the DBF4 pre-mRNA indicated an enrichment in DBF4 signal after 3.5 h and 21 h. L: ladder, -RT: samples lacking reverse transcriptase ruling out DNA contamination
C. Similar to DBF4 pre-mRNA, primer extension revealed an enrichment of signal for the DBF4$_{ups}$ transcripts after 3.5 and 21 h treatment. L: ladder

from HeLa cells using a Click-iT RNA Nascent RNA Capture Kit (Invitrogen; Burlington, Ontario, Canada) were analyzed for their degradation over time to determine the half-life of the *DBF4*$_{ups}$ transcripts (**Figure 8**). Using the kit, RNA was tagged with ethynyl uridine (EU), a uracil analogue, which incorporates into the RNA of exposed cells as the RNA is being synthesized. EU-labeled RNA can be selectively isolated using a Click-iT RNA Nascent RNA Capture Kit. EU was first added to the cell medium to allow cells the time to take up EU and incorporate it into the RNA being synthesized. After a period of 12 h or 24 h, EU was removed from the culture medium by substituting it with fresh medium. EU-labeled RNA may degrade over time, from which the half-life of transcripts can be determined. Transcripts containing EU at T_0, the time at which EU was removed from the medium, were compared against the incubation timepoints after T_0. RNA that is less stable will deplete from the pool of EU-labeled at a faster rate. The half-life of the *DBF4*$_{ups}$ transcript was calculated to be approximately 3.7 h while the half-lives of the low-turnover control RN18S and the high-turnover control UBC were calculated to be 14 h and 4.4 h respectively. The high turnover of *DBF4*$_{ups}$ suggests that the reduction in expression level would have been observed within the 20 h treatment time frame if RNAP I, II or III were responsible for the transcription of *DBF4*$_{ups}$ RNA.

3.4 Knockdown of MtRPOL using siRNA reduces the expression of *DBF4*$_{ups}$ RNA

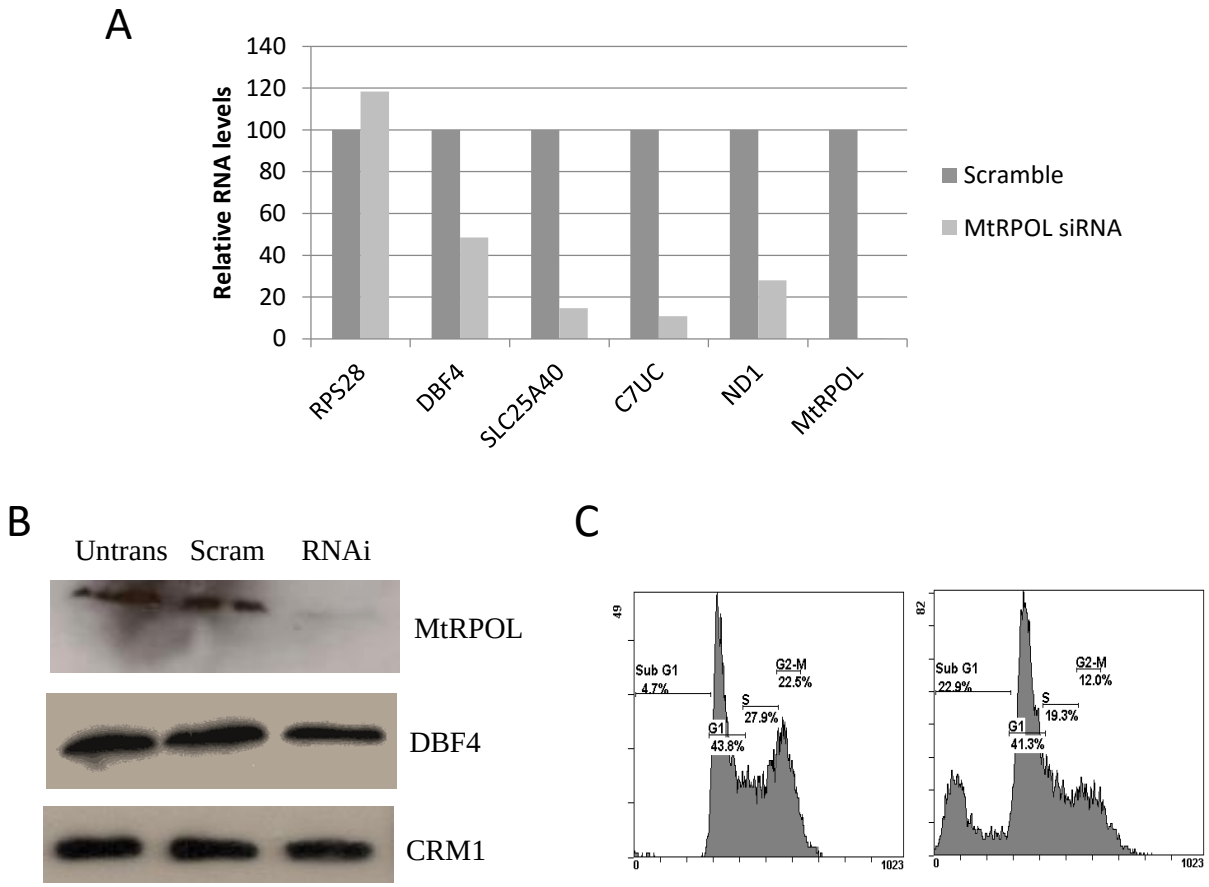
While RNAP I, II and III were able to be directly inhibited with inhibitors that compromised each of the RNA polymerases ability to transcribe their targets, there are no well-known inhibitors of MtRPOL. While MtRPOL activity inhibition has been recently reported to be an off-target effect of antiviral nucleosides (AVRNs) (Arnold et al. 2012), we have not been able to inhibit MtRPOL using AVRNs. I therefore attempted to ablate MtRPOL activity using

Figure 8. The *DBF4$_{ups}$* transcripts have a relatively short half-life



The rate of RNA degradation was monitored on labeled RNA using a Click-iT RNA Capture Kit. HeLa cells were incubated in medium containing ethynyl uridine (EU) for 24h. Levels of labeled RNA from samples collected 0 h after release into fresh medium in the absence of EU (EU-) and used as the pre-degradation baseline. The fold decrease between 0 h and 12 h (**A**) or 0 h and 24 h (**B**) post-release into EU- medium were measured. The differences in expression levels compared to the 0 h time point were measured using qPCR with primers for the stable 18S rRNA (low turnover control), ubiquitin ligase C (UBC, a transcript that has a relatively short half-life) and *DBF4$_{ups}$*. While the labeled 18S rRNA showed little loss in both time points, the levels of *DBF4$_{ups}$* transcripts decreased substantially, suggesting that the *DBF4$_{ups}$* half-life is much shorter. standard deviation is indicated in **A**.

Figure 9. siRNA-mediated MtRPOL knockdown may affect DBF4ups expression levels



A. Expression levels of multiple gene transcripts were measured using RT-qPCR for HEK293T cells transfected with MtRPOL siRNA for 72 h. The RNA coding for the ribosomal protein RPS28 was used as a control to assure even concentrations between samples. Expression level is normalized to 100% expression being non-treated, treated cells were also normalized to differences in RPS28.

B. Western blot with an antibody directed towards MtRPOL confirms the knockdown of the protein at 72 h post-siRNA transfection, note that DBF4 and CRM1 (loading control) levels remain unaffected. Untrans: untransfected control, Scram: transfected with a random siRNA, RNAi: MtRPOL-directed RNAi treated cells.

C. Flow cytometry 72 h after transfection indicates a reduction in the G2/M peak as well as sub-G1 accumulation indicating cell death.

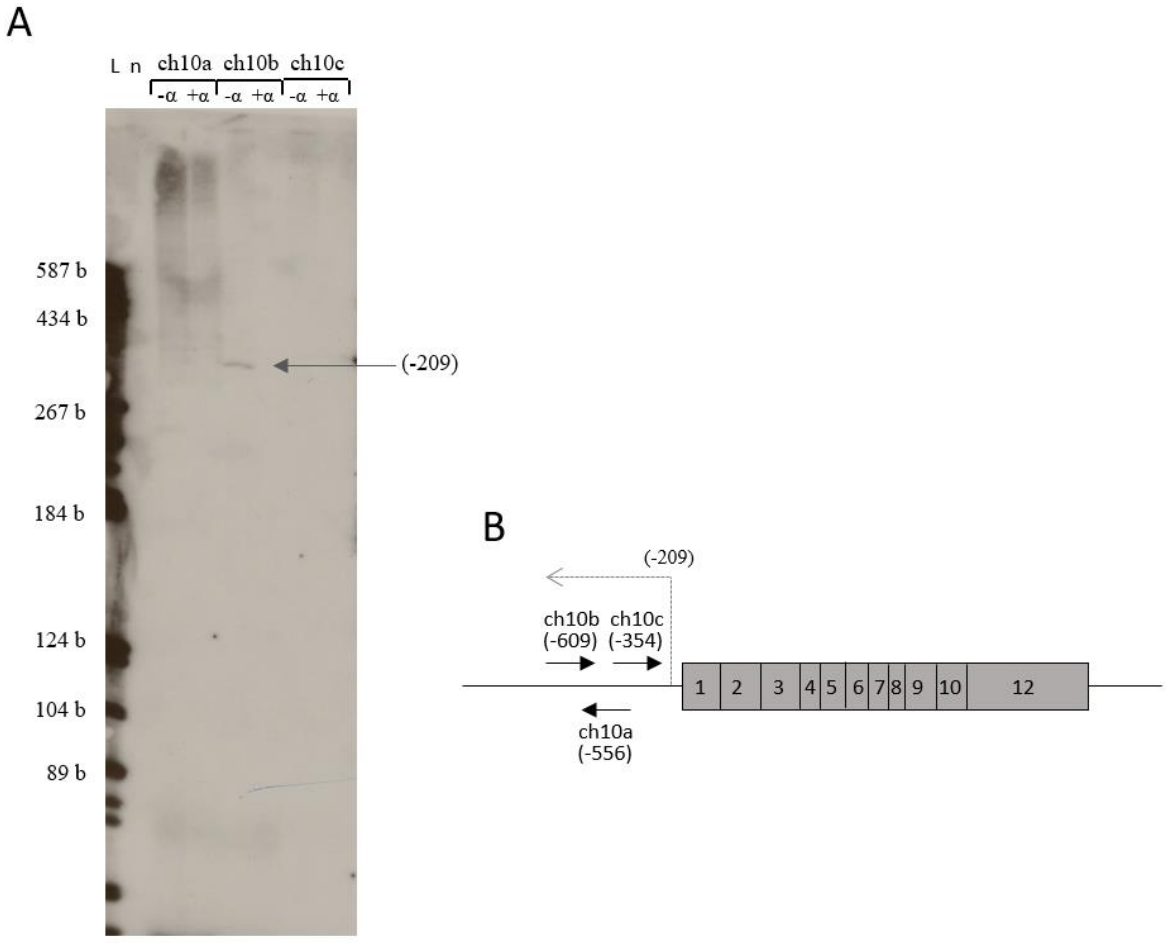
RNAi. By 72 h of inhibition with siRNA (**Figure 9**), the levels of both MtRPOL and *DBF4*$_{ups}$ were reduced by 72h post-treatment with siRNA. This suggests that *DBF4*$_{ups}$ may be transcribed by MtRPOL (**figure 9A**). Unexpectedly, however, the levels of *DBF4* (primer C7UC) and *SLC25a40* were also substantially downregulated (**Figure 9A**). *DBF4*$_{\text{ups}}$ expression was reduced even more drastically than the mitochondrial ND1 gene, which could be due to the considerably short half-life for *DBF*$_{\text{ups}}$ determined previously (**Figure 8**). Alternatively, this result could indicate that the expression of DBF4 and SLC25a40 were directly affected by the expression of *DBF4*$_{ups}$, perhaps by regulating the chromatin structure.

3.5 No α-amanitin-resistant lncRNAs were found at the *DBF4* pseudogene

Although the regions immediately upstream of the *DBF4* transcribed region and the *DBF4* pseudogene are unique for each chromosome, there is always the possibility that some of the primers used that target within the first exon of DBF4 could also be picking up signals from the *DBF4* pseudogene, obscuring the results. While most of the primers used target within the *DBF4* promoter region, the DIG-labeled primers pA and pC target sequences within the first exon of *DBF4* and have high sequence similarity with the pseudogene. To assess whether these DIG-label primers may be picking up signals from the *DBF4* pseudogene transcripts, primer extension was carried out on total RNA isolated from HeLa cells using three DIG-labeled primers targeting immediately upstream of the *DBF4* pseudogene (**Figure 10**).

A product of approximately 300 bases in size was detected using the DIG-labeled primer ch10b (**Figure 10A**), which would effectively place its TSS at -209 bases from the pseudogene's "ATG" sequence in the antisense direction. A faint smear of bands were also obtained with ch10a between -850 and beyond -1200 base in the sense direction. Unlike the *DBF4*$_{ups}$ and *DBF4*$_{as}$

Figure 10. The DBF4 pseudogene has very low transcription activity



A. Primer extension was used to detect lncRNA TSSs immediately upstream of the *DBF4* pseudogene locus in the absence (-α) or presence (+α) of α-amanitin. Primer ch10a was designed to target potential transcripts in a sense direction to the *DBF4* pseudogene while primers ch10b and ch10c target potential lncRNAs that are antisense to the pseudogene. Very little signal was produced in this region with only one strongly defined band (-209 from ATG sequence of the pseudogene). However, the observed signals were sensitive to α-amanitin, unlike the *DBF4* promoter region. L: molecular weight marker, n: no template control

B. A map of the *DBF4* pseudogene on chromosome 10. The regions corresponding to each individual exon of the original gene have been delimited and numbered with exon 11 notably missing. The positions of the DIG-labelled primers used in A have also been indicated with the position of the 5' end of the primer indicated in relation of the ATG sequence of the pseudogene with the position of the -209 band indicated in **A**.

lncRNAs detected previously, all of the signals observed using the primers directed to the upstream region of the *DBF4* pseudogene were sensitive to α-amanitin treatment. Furthermore, none of these signals corresponded to the TSSs of the lncRNAs previously detected using primers pA and pC. We can therefore rule out that the results from the previous experiment are not misrepresented due to the presence of the *DBF4* pseudogene.

4.0 Discussion

4.1 LncRNAs of the *DBF4* locus

The results from primer extension indicate that the multiple $DBF4_{ups}$ transcripts are at a minimum length of 200 to 450 bases long. Although the shortest lncRNA was under 100 bases in size, the Qiagen RNeasy mini kit isolates RNA that are over 200 bases long indicating the template from which it was extended is likely longer than 200 bases. The results of primer extension only reveal that there is a product of a certain size originating from the site and the direction that the DIG-labeled primer is targeting. This method does not yield a definitive sequence and, therefore, does not account for possible transcript splicing. The longest lncRNA indicated by the DIG-labeled primer pE (which targets -402) is approximately 450 bases long, indicating an unspliced sequence would end at -852.

Splicing of RNA is most prominent in RNAP II-dependent transcripts and is accomplished with the help of a protein-RNA complex known as the spliceosome (Martins et al. 2011). Other forms of splicing are present but are very rare and have only been identified in a handful of genes. Rybozymes in plant organelles, simple eukaryotes and bacteria have self-splicing introns group II splicing with splice sites similar to those spliced by the spliceosome (Brown et al. 2014), however, these have not been reported in mammals. Yeast tRNAs have also been reported to undergo an unrelated form of protein-mediated splicing (Trotta et al. 1997). While self-splicing introns have been reported in yeast mitochondrial RNA, human mitochondrial RNA are completely devoid of intronic regions and thus no splicing may take place. To assess potential splicing of the DBF4 lncRNAs, the NNSPLICE human splice site prediction tool was used for the -402 to -852 sequence. A possible splicing acceptor site (score 0.85) was predicted at position -666 indicating that there is a possibility that the longest of the

detected RNAs could be spliced. Two primer sets used in PCR, Prom 7B and Prom C7UC, also successfully detected the α-amanitin-resistant transcripts in this region and were able to amplify regions -625 to -675 and -640 to -703, respectively. Detection of the lncRNAs using these PCR primers suggests that the *DBF4*$_{ups}$ transcripts are not spliced because the -666 acceptor site is located within the amplified PCR product. The *DBF4*$_{as}$ transcripts show 3 putative acceptor sites on NNSPLICE analysis. These potential splice sites have not been confirmed, however, and further research is needed to verify if these sites are functional splicing sites.

Attempts to determine the remaining 3' end of the *DBF4*$_{ups}$ sequence have so far been unsuccessful. Expressed sequence tags (ESTs) databases consist of randomly sequenced incomplete RNA fragments. ESTs are generally thought to be RNAP II transcribed mRNA as they are sequenced by creating cDNAs from their poly-adenosine tails (Adams et al. 1991). Interestingly, the *DBF4*$_{ups}$ transcripts were also present in samples isolated for polyadenylated RNA (unpublished data) indicating that these transcripts have a poly-A tail and may also be detected in EST libraries. While a few ESTs indicates there is some noise between the *DBF4* and *SLC25a40* genes, very little information on the full sequence of the *DBF4*$_{ups}$ transcripts are present in online sequence browsers. Technical limitations make it difficult to perform genome-wide studies to search for RNAP I and III transcripts as many of the genome-wide sequencing techniques rely on targeting capped transcripts to acquire the 5' ends of the RNA to either target specific RNAP II transcripts or to filter out all of the RNA that is currently undergoing RNAse degradation in the sample (Kodzius et al. 2006). Other techniques simply sequence all RNA indiscriminately making it difficult to differentiate overlapping transcripts (Martin & Wang 2011). The capping enzyme complex (CEC) which is responsible for capping nascent transcripts associates uniquely with RNA polymerase II (Cho et al. 1997). If the *DBF4*$_{ups}$ and *DBF4*$_{as}$

transcripts are not transcribed by RNA polymerase II, they are not likely capped and therefore may evade most mass sequencing techniques

As previously noted, the *DBF4*$_{ups}$ transcripts were found to have poly-A tails. In mammals, only two types of transcripts typically contain poly-A tails: long poly-A tails are found on RNA transcribed by RNAP II and shorter tails on those transcribed by the mitochondrial RNAP (Nagaike et al. 2008). Additionally, the TRAMP complex has been reported to polyadenylate ncRNA with a short series of adenosines and marking them for degradation, although it is not known if this process is also present in human cells (Hamill et al. 2010; Callahan & Butler 2010).

A reduction in the expression of the RNAP II-transcribed SLC25a40 was an unexpected result that occurred during siRNA-mediated inhibition of MtRPOL (**Figure 9A**). Flow cytometry analysis of these samples also indicated a significant level of cell death (**Figure 9C**). It is possible that knockdown of MtRPOL is too toxic to the cell and the observed changes are side effects of dying cells in which RNA begins degrading rapidly. Alternatively, because *SLC25a40* itself codes for a mitochondrial solute carrier protein, it is possible that knockdown of MtRPOL caused feedback from the mitochondria for cells to reduce transcription of some mitochondria-related proteins. While the general trend has been that the cells that were not expressing MtRPOL had reduced *DBF4*$_{ups}$ expression.

While current data on the *DBF4* lncRNAs support the possibility that MtRPOL transcribes these transcripts in the nucleus, I have been unable to verify the existence of the reported nuclear form of MtRPOL as described previously (Kravchenko et al. 2005). A recent study also claimed that MtRPOL does not transcribe nuclear genes or localize to the nucleus (Kühl et al. 2014), indicating that further research is necessary to determine the elusive source of

the transcription of *DBF4* lncRNAs. Furthermore, even if MtRPOL located to the nucleus, it is still unclear which transcription factors would be required for MtRPOL to interact with chromosomal DNA. It is not known if MtRPOL would be able to navigate around histones as these proteins are absent in mitochondrial DNA (Alexeyev et al. 2013). In the mitochondria, MtRPOL requires transcription factors A and B2 (TFAM and TFB2M) to be able to load onto DNA. TFB2M has not been reported to localize in the nucleus while one study reported TFAM can be localized in the nucleus of human prostate PC3 cells (Pastukh et al. 2007). Alternatively, a nuclear MtRPOL may be interacting with proteins normally located in the nucleus.

4.2 Features of the *DBF4* promoter chromatin landscape

Transcription of the $DBF4_{ups}$ lncRNAs begins at many sites close to the two replication initiation zones where the ORCs bind (**Figure 1**). While some transcripts have been shown to initiate further downstream of the *DBF4* promoter region, these transcripts showed sensitivity to RNAP II inhibition, suggesting that they are not part of the same set of transcripts. Interestingly, asymmetric replication at each initiation zone proceeds in the same direction as the lncRNAs that begin transcription in each zone.

Considering their proximity to replication origins, one of the possible functions of these lncRNAs may be their involvement in the regulation of physical access of the replication proteins to the location of the replication origins. Crosstalk between transcription and replication has been previously reported as a large set of replication origins are co-localized to active gene loci (Méchali 2010). The firing of early replication origins is correlated with origins located at heavily transcribed regions, while late-firing origins are often located in regions with little to no transcription activity (Schwaiger & Schübeler 2006). While it is not known exactly what

determines the timing of a replication origin, it has been suggested that open chromatin configurations can be affected by the transcription machinery in these areas by opening up the region of densely packed heterochromatin to allow for easier access for replication proteins (Rhind 2006). It is possible that transcription of the *DBF4* lncRNAs contribute to opening up the chromatin to allow for the ORCs to bind. A previous study suggests that ORC binding could be RNA dependent. Therefore, the *DBF4* lncRNAs could alternatively play a role in stabilizing ORCs (Norseen et al. 2008). Further research is required to be able to determine the exact mechanism by which lncRNAs at *DBF4* may be involved in the regulation of the *DBF4* origin.

While *DBF4* mRNA levels fluctuate throughout the S phase, the $DBF4_{ups}$ transcripts remain constant (**Fig. 5**). Many genes with neighboring lncRNAs are reported to coordinate transcription with each other, possibly due to overlapping transcription and competition for transcription factors (Hiratani et al. 2009). The decrease in $DBF4_{ups}$, *DBF4* and *SLC25a40* in **Figure 9** could suggest that the absence of $DBF4_{ups}$ may have altered the chromatin landscape at the *DBF4* promoter locus and influenced *DBF4* and *SLC25A40*. The $DBF4_{ups}$ transcription appears to be regulated by a novel mechanism potentially involving MtRPOL as inhibiting all known nuclear RNA polymerases has no effect on DBF_{ups} transcript levels. The lack of consensus on whether or not the mitochondrial RNA polymerase is able to locate to the nucleus however stresses a need for more rigorous experiments to determine if MtRPOL is responsible for transcribing the *DBF4* lncRNAs.

Conclusion

In this thesis, I identify a novel lncRNAs located near the *DBF4* promoter region that does not appear to behave like other known nuclear RNAs in terms of transcription. The first objective was to determine if the *DBF4* lncRNAs were transcriptionally different from the *DBF4* gene with which they overlap. The DBF_{ups} transcripts are differentially transcribed from the *DBF4* gene throughout the cell cycle and, unlike *DBF4*, are resistant to α-amanitin treatment. $DBF4_{ups}$ transcripts also have a half-life of approximately 3.7 h.

It is still unclear as to which polymerase is responsible for transcribing the *DBF4* promoter region to generate the lncRNAs. Inhibiting RNA polymerases I, II and III yielded no apparent reduction in expression and in some cases the lncRNAs signal became stronger. MtRPOL is currently the strongest candidate for transcribing the *DBF4* lncRNAs, although there are conflicting reports on whether or not it localizes to in the nucleus (Kühl et al. 2014; Lee et al. 2011; Kravchenko et al. 2005). With the doubt surrounding whether or not MtRPOL, or a truncated form of the polymerase, locates to the nucleus, further research is required to confirm if MtRPOL is transcribing the lncRNAs discussed in this thesis.

Lastly, in an attempt to determine if reduced transcription of the $DBF4_{ups}$ transcripts had an effect on DBF4, MtRPOL was knocked down. While knockdown of MtRPOL Yielded significantly reduced levels of $DBF4_{ups}$ transcripts, it also similarly affected the *SLC25a40* gene and lowered levels of DBF4 as well. It is unclear if the reduction is a result of reduction of DB4ups transcription in the area making the promoter region for these two genes less accessible or if it was due to the toxicity invoked on cells as a consequence of knocking down MtRPOL. Finding a siRNA that is able to specifically knock down the *DBF4* lncRNAs may yield more specific results.

The positioning of these lncRNAs make them an attractive target to investigate the complex relations between origin activity, gene transcription and lncRNA transcription at a single genomic locus. Further study on this aspect can open up an exciting new opportunity of unravelling a novel mechanism of DNA replication by lncRNA transcription in the context of MtRPOL-mediated transcription.

References

Abdurashidova, G. et al., 2000. Start sites of bidirectional DNA synthesis at the human lamin B2 origin. *Science (New York, N.Y.)*, 287(5460), pp.2023–2026.

Adams, M.D. et al., 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252(5013), pp.1651–6.

Alexeyev, M. et al., 2013. The maintenance of mitochondrial DNA integrity--critical analysis and update. *Cold Spring Harbor perspectives in biology*, 5(5).

Arias, E.E. & Walter, J.C., 2007. Strength in numbers: Preventing rereplication via multiple mechanisms in eukaryotic cells. *Genes and Development*, 21(5), pp.497–518.

Arnold, J.J. et al., 2012. Human mitochondrial RNA polymerase: Structure–function, mechanism and inhibition. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1819(9-10), pp.948–960. Available at: http://dx.doi.org/10.1016/j.bbagrm.2012.04.002.

Ballabeni, A. et al., 2004. Human geminin promotes pre-RC formation and DNA replication by stabilizing CDT1 in mitosis. *The EMBO journal*, 23(15), pp.3122–3132.

Bell, S.P. & Dutta, A., 2002. DNA replication in eukaryotic cells. *Annual review of biochemistry*, 71, pp.333–374.

Bielinsky, A.K. et al., 2001. Origin recognition complex binding to a metazoan replication origin. *Current Biology*, 11(18), pp.1427–1431.

Blagosklonny, M. V., 2006. Cell senescence: Hypertrophic arrest beyond the restriction point. *Journal of Cellular Physiology*, 209(3), pp.592–597.

Bostock, C.J., Prescott, D.M. & Kirkpatrick, J.B., 1971. An evaluation of the double thymidine block for synchronizing mammalian cells at the G1-S border. *Experimental cell research*, 68(1), pp.163–168.

Brown, G.G., Colas des Francs-Small, C. & Ostersetzer-Biran, O., 2014. Group II intron splicing factors in plant mitochondria. *Frontiers in plant science*, 5(February), p.35. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3927076&tool=pmcentrez&rendertype=abstract.

Burhans, W.C. et al., 1990. Identification of an origin of bidirectional DNA replication in mammalian chromosomes. *Cell*, 62(5), pp.955–965.

Callahan, K.P. & Butler, J.S., 2010. TRAMP complex enhances RNA degradation by the nuclear exosome component Rrp6. *Journal of Biological Chemistry*, 285(6), pp.3540–3547.

Cho, E.J. et al., 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes and Development*, 11(24), pp.3319–3326.

Chow, J.C. et al., 2005. Silencing of the mammalian X chromosome. *Annual review of genomics and human genetics*, 6, pp.69–92.

Cordaux, R. & Batzer, M.A., 2009. The impact of retrotransposons on human genome evolution. *Nature reviews. Genetics*, 10(10), pp.691–703.

Dellino, G.I. et al., 2013. Genome-wide mapping of human DNA-replication origins: Levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Research*, 23(1), pp.1–11.

DePamphilis, M.L., 2003. The "ORC cycle": A novel pathway for regulating eukaryotic DNA replication. *Gene*, 310(1-2), pp.1–15.

Drygin, D., Rice, W.G. & Grummt, I., 2010. The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. *Annual review of pharmacology and toxicology*, 50, pp.131–156.

Eggermont, J. & Proudfoot, N.J., 1993. Poly(A) signals and transcriptional pause sites combine to prevent interference between RNA polymerase II promoters. *The EMBO journal*, 12(6), pp.2539–2548.

Elledge, S.J., 1996. Cell cycle checkpoints: preventing an identity crisis. *Science (New York, N.Y.)*, 274(5293), pp.1664–1672.

Elmore, S., 2007. Apoptosis: a review of programmed cell death. *Toxicologic pathology*, 35(4), pp.495–516.

Gilbert, D.M., 2001. Making Sense of Eukaryotic DNA Replication Origins. *Science*, 294(5540), pp.96–100.

Giri, S. & Prasanth, S.G., 2012. Replicating and transcribing on twisted roads of chromatin. *Briefings in Functional Genomics*, 11(3), pp.188–204.

Goodrich, J.A. & Kugel, J.F., 2006. Non-coding-RNA regulators of RNA polymerase II transcription. *Nat Rev Mol Cell Biol*, 7(8), pp.612–616. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16723972\nhttp://www.nature.com/nrm/journal/v7/n8/pdf/nrm1946.pdf.

Hahn, S., 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nature structural & molecular biology*, 11(5), pp.394–403.

Hamill, S., Wolin, S.L. & Reinisch, K.M., 2010. Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. *Proceedings of the National Academy of Sciences of the United States of America*, 107(34), pp.15045–15050.

Hiratani, I. et al., 2009. Replication timing and transcriptional control: beyond cause and effect-part II. *Current Opinion in Genetics and Development*, 19(2), pp.142–149.

Jiang, W. et al., 1999. Mammalian Cdc7-Dbf4 protein kinase complex is essential for initiation of DNA replication. *The EMBO journal*, 18(20), pp.5703–5713.

Johnson, D.G. & Walker, C.L., 1999. Cyclins and cell cycle checkpoints. *Annual review of pharmacology and toxicology*, 39, pp.295–312.

Kastan, M.B. & Bartek, J., 2004. Cell-cycle checkpoints and cancer. *Nature*, 432(7015), pp.316–323.

Kodzius, R. et al., 2006. CAGE: cap analysis of gene expression. *Nature methods*, 3(3), pp.211–222.

Kravchenko, J.E. et al., 2005. Transcription of mammalian messenger RNAs by a nuclear RNA polymerase of mitochondrial origin. *Nature*, 436(7051), pp.735–739.

Kühl, I. et al., 2014. POLRMT does not transcribe nuclear genes. *Nature*, 514(7521), pp.E7–E11. Available at: http://www.nature.com/doifinder/10.1038/nature13690.

Kumagai, H. et al., 1999. A novel growth- and cell cycle-regulated protein, ASK, activates human Cdc7-related kinase and is essential for G1/S transition in mammalian cells. *Molecular and cellular biology*, 19(7), pp.5083–5095.

Lee, H. & Romero, J., 2012. Origin of DNA replication at the human lamin B2 locus: OBR or ABR? *Cell Cycle*, 11(22), pp.4281–4283.

Lee, J.T., 2012. Epigenetic regulation by long noncoding RNAs. *Science*, 338(6113), pp.1435–1439. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23239728\nhttp://www.sciencemag.org/content/338/6113/1435.full.pdf.

Lee, Y.L., Chiao, C.H. & Hsu, M.T., 2011. Transcription of muscle actin genes by a nuclear form of mitochondrial RNA polymerase. *PLoS ONE*, 6(7).

Luco, R.F. & Misteli, T., 2011. More than a splicing code: integrating the role of RNA, chromatin and non-coding RNA in alternative splicing regulation. *Current opinion in genetics & development*, 21(4), pp.366–372. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=21497503&retmode=ref&cmd=prlinks\npapers3://publication/doi/10.1016/j.gde.2011.03.004.

Ma, L., Bajic, V.B. & Zhang, Z., 2013. On the classification of long non-coding RNAs. *RNA biology*, 10(6), pp.925–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23696037.

Magdalou, I. et al., 2014. The causes of replication stress and their consequences on genome stability and cell fate. *Seminars in Cell and Developmental Biology*, 30(30), pp.154–164.

Martin, J. a & Wang, Z., 2011. Next-generation transcriptome assembly. *Nature reviews. Genetics*, 12(10), pp.671–682. Available at: http://dx.doi.org/10.1038/nrg3068.

Martins, S.B. et al., 2011. Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3′ end of human genes. *Nature Structural & Molecular Biology*, 18(10), pp.1115–1123. Available at: http://dx.doi.org/10.1038/nsmb.2124.

Méchali, M., 2010. Eukaryotic DNA replication origins: many choices for appropriate answers. *Nature reviews. Molecular cell biology*, 11(10), pp.728–738. Available at: http://dx.doi.org/10.1038/nrm2976.

Mercer, T.R., Dinger, M.E. & Mattick, J.S., 2009. Long non-coding RNAs: insights into functions. *Nature reviews. Genetics*, 10(3), pp.155–159.

Nagaike, T., Suzuki, T. & Ueda, T., 2008. Polyadenylation in mammalian mitochondria: Insights from recent studies. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1779(4), pp.266–269.

Norseen, J. et al., 2008. RNA-dependent recruitment of the origin recognition complex. *The EMBO journal*, 27(22), pp.3024–3035.

Pastukh, V. et al., 2007. Human mitochondrial transcription factor A possesses multiple subcellular targeting signals. *FEBS Journal*, 274(24), pp.6488–6499.

Phatnani, H.P. & Greenleaf, A.L., 2006. Phosphorylation and functions of the RNA polymerase II CTD. *Genes and Development*, 20(21), pp.2922–2936.

Rhind, N., 2006. DNA replication timing: random thoughts about origin firing. *Nature cell biology*, 8(12), pp.1313–1316.

Romero, J., 2008. *Identification and Detailed Characterization of a Novel Origin of DNA Replication in the Trancriptional Promoter of the Human DBF4 Gene*.

Romero, J. & Lee, H., 2008a. Asymmetric bidirectional replication at the human DBF4 origin. *Nature structural & molecular biology*, 15(7), pp.722–729.

Romero, J. & Lee, H., 2008b. One-way PCR-based mapping of a replication initiation point (RIP). *Nature protocols*, 3(11), pp.1729–1735.

Schafer, K.A., 1998. The cell cycle: a review. *Veterinary pathology*, 35(6), pp.461–478.

Schneider, R. & Grosschedl, R., 2007. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes and Development*, 21(23), pp.3027–3043.

Schwaiger, M. & Schübeler, D., 2006. A question of timing: Emerging links between transcription and replication. *Current Opinion in Genetics and Development*, 16(2), pp.177–183.

Slomovic, S. et al., 2010. Addition of poly(A) and poly(A)-rich tails during RNA degradation in the cytoplasm of human cells. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16), pp.7407–7412.

Tijsterman, M. & Plasterk, R.H.A., 2004. Dicers at RISC: The mechanism of RNAi. *Cell*, 117(1), pp.1–3.

Trotta, C.R. et al., 1997. The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell*, 89(6), pp.849–858.

Vogelauer, M. et al., 2002. Histone acetylation regulates the time of replication origin firing. *Molecular Cell*, 10(5), pp.1223–1233.

Weinreich, M. & Stillman, B., 1999. Cdc7p-Dbf4p kinase binds to chromatin during S phase and is regulated by both the APC and the RAD53 checkpoint pathway. *EMBO Journal*, 18(19), pp.5334–5346.

White, R.J., 2011. Transcription by RNA polymerase III: more complex than we thought. *Nature*

reviews. Genetics, 12(7), pp.459–463.

Will, C.L. & Lührmann, R., 2011. Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology*, 3(7), pp.1–2.