Supporting Online Material for:

PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters

Pascal Preker[1], Kristina Almvig[1], Marianne S. Christensen[1], Eivind Valen[2], Cristophe K. Mapendano[1], Albin Sandelin[2] and Torben Heick Jensen[1,*]

[1]Centre for mRNP Biogenesis and Metabolism, Department of Molecular Biology, C.F. Møllers Allé, bldg. 1130, Aarhus University, 8000 Aarhus, Denmark and [2]The Bioinformatics Centre, Dept. of Biology and the Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark

*corresponding author; email: thj@mb.au.dk; phone: (+45) 6020 2705; fax: (+45) 8619 6500

This Supporting Material contains:

Online Material and Methods

Four Supplemental Figures

Two Supplemental Tables

Online Material and Methods

ChIP assays

For CTD-Ser2P and CTD-Ser5P experiments, phoshatase inhibitors (see main text) were added to the dilution buffer (16.7 mM Tris/HCl pH 8.1, 1.2 mM EDTA, 167 mM NaCl, 0.01% SDS, 1.1% Triton X-100). Antibodies were added to 2 ml of a 10 fold dilution of NCE in dilution buffer and incubated overnight at 4°C on a rotator in the presence of 30 μl protein A sepharose CL-4B beads (GE Healthcare), pre-blocked with 1 mg/ml BSA and 0.2 mg/ml salmon sperm DNA. Next day, beads were washed once in each of the following buffers: low-salt wash buffer (20 mM Tris/HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), high-salt wash buffer (same as low-salt wash buffer but with 500mM NaCl), LiCl wash buffer (10 mM Tris/HCl pH 8.1, 250 mM LiCl, 1mM EDTA, 1% NP-40, 1% natriumdeoxycholat) and twice in TE buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA). IP'ed complexes were eluted for 30 min at room temperature in 200 μl elution buffer (1% SDS, 100 mM $NaHCO_3$). Cross-linking was reversed by incubating eluate and input samples at 65°C overnight in elution buffer containing 125 mM NaCl and 25 μg of DNase-free RNase A. EDTA and Tris/HCl ph7.0 were added to final concentrations of 20 and 80 mM, respectively, and samples were treated with 100 μg Proteinase K for 3 hr at 45°C. Finally, the DNA was recovered using a QIAquick PCR purification kit (Qiagen). Eluates (100 μl in 10 mM Tris/HCl pH 8.5) were diluted four fold with H_2O. Six μl (1.5%) of each sample were used per qPCR reaction as described above. All samples were analyzed in triplicate, and results are expressed as the mean percent of input after subtraction of background values ("No Antibody Control").

Supplemental Figure Legends

Supplemental Figure 1: PROMPTs are enriched in the cell nucleus

(a) Equal amounts (5 µg) of nuclear ('Nuc') or cytoplasmic ('Cyt') RNA from fractionated HeLa cells were resolved on a 1% agarose gel and stained with ethidium bromide. Positions of pre-, 28S- and 18S-ribosomal RNA species as well as DNA marker fragments ('M') are indicated.

(b) Validation of HeLa nuclear and cytoplasmic fractions. Total and subcellular protein fractions were isolated in parallel with RNA preparations shown in (a) and subjected to western blotting analysis with

antibodies towards a cytoplasmic (tubulin, upper panel) and a nuclear (U1-70K, lower panel) marker protein at dilutions of 1:2,500 and 1:5,000, respectively. The migration of protein size markers (kDa) is indicated.

(c) Nuclear (N) and cytoplasmic (C) fractions were analyzed by RT-PCR (see Figure 1D) for contents of the indicated RNAs. Products of 40 cycles of amplification are visualized on a 2% agarose gel stained with ethidium bromide and DNA size markers indicated on the left.

Supplemental Figure 2: 5' RACE analyses of DEPDC5 and SPHK1 PROMPT regions

See legend to Fig. 3Dfor details.

Supplemental Figure 3: Similar relative levels of CTD-Ser5P in PROMPT and gene regions

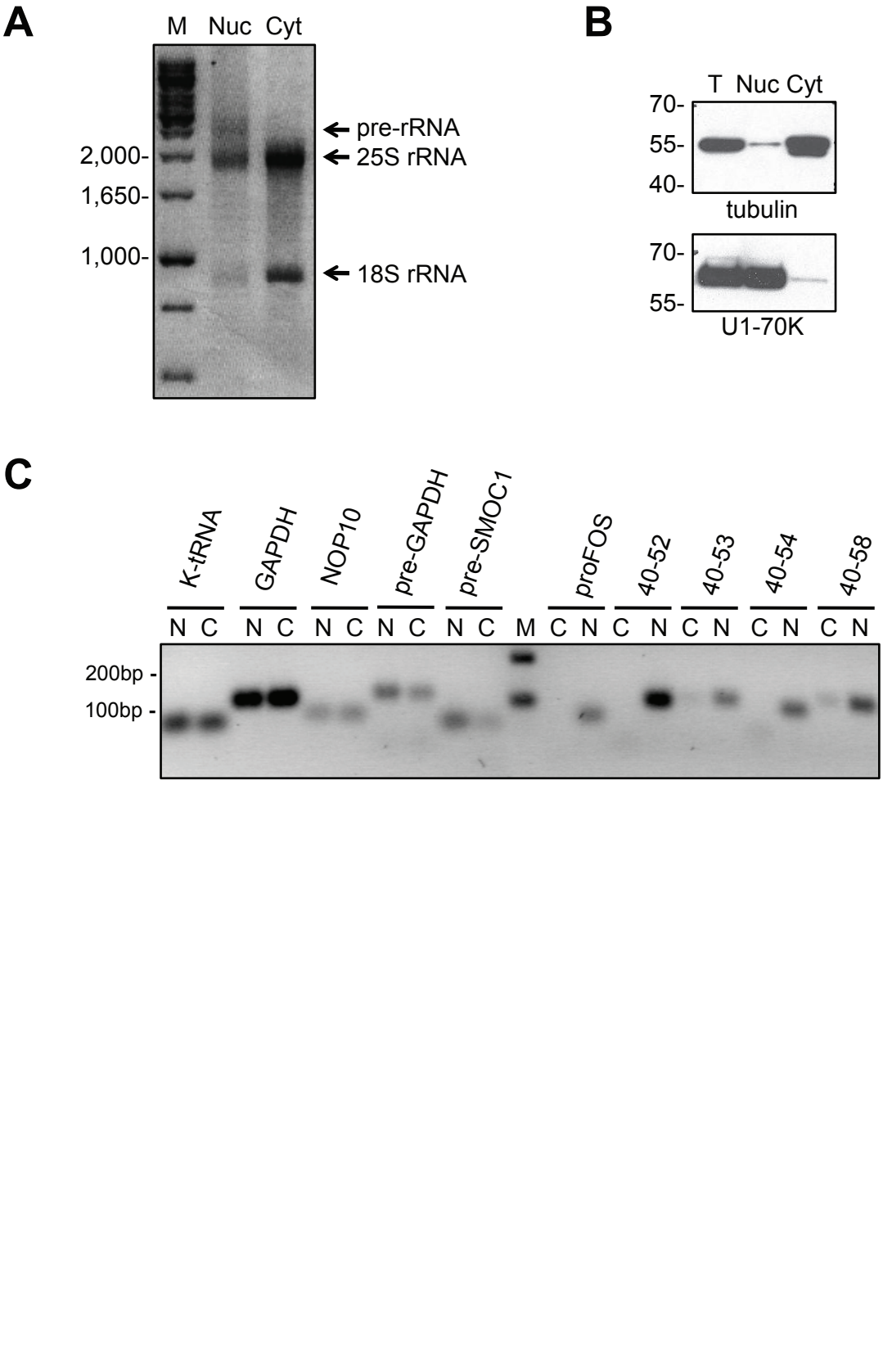
(a) ChIP analysis of the MYC and CCND1 gene and PROMPT regions using an antibody specific for CTD-Ser5P. Values show the mean percent of input recovered in the IP as determined by qPCR. Error bars indicate means of standard deviation of triplicate technical replica of a representative experiment.

Supplemental Figure 4: Location of RNAPIII genes near unspliced expressed sequence tags (ESTs) and convergently transcribed RNAPII genes

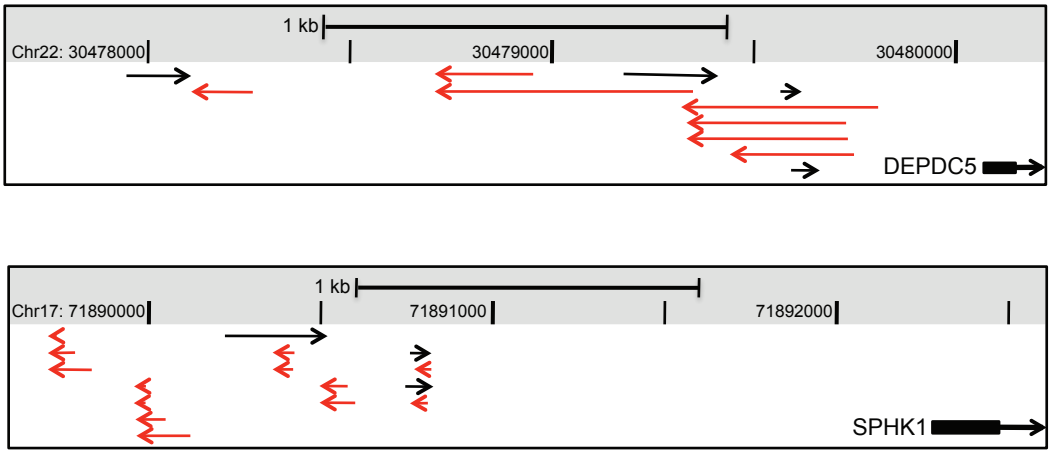
(a) RNAPIII PROMPT regions overlapping "orphan" ESTs. UCSC genome browser view of hY5 (upper panel) and SRP (lower panel) RNAPIII-genes and their upstream regions. The locations of amplicons used for RT-qPCR and ChIP analysis in Figs. 4 and 5 are represented by blue boxes above the ESTs (black, labeled by accession numbers). The RNAPIII genes are depicted in brown with an embedded white arrow indicating the direction of transcription. Chromosomal positions (hg18) are on top.

(b) RNAPIII transcripts located in the PROMPT regions of RNAPII genes. UCSC genome browser view of three examples of RNAPIII genes whose transcription originate upstream of, and in antisense orientation to, protein-coding genes (blue tracks). The labeling convention follows (a) except that G/C percentage (black/grey) and vertebrate conservation (green) are shown as separate tracks at the top and bottom, respectively, of each panel.

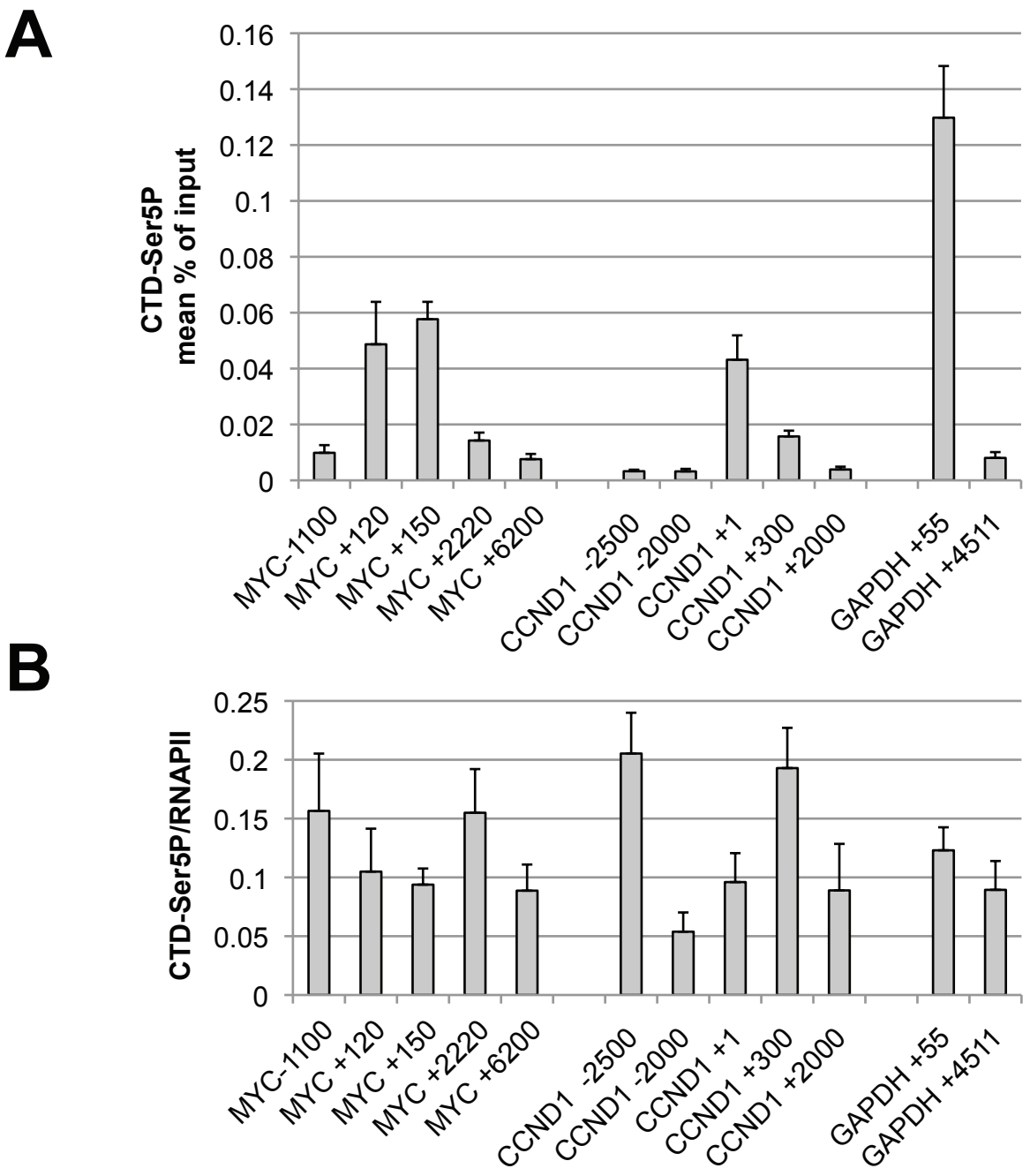
Preker et al., Sup. Fig. 1



Preker et al., Sup. Fig. 2

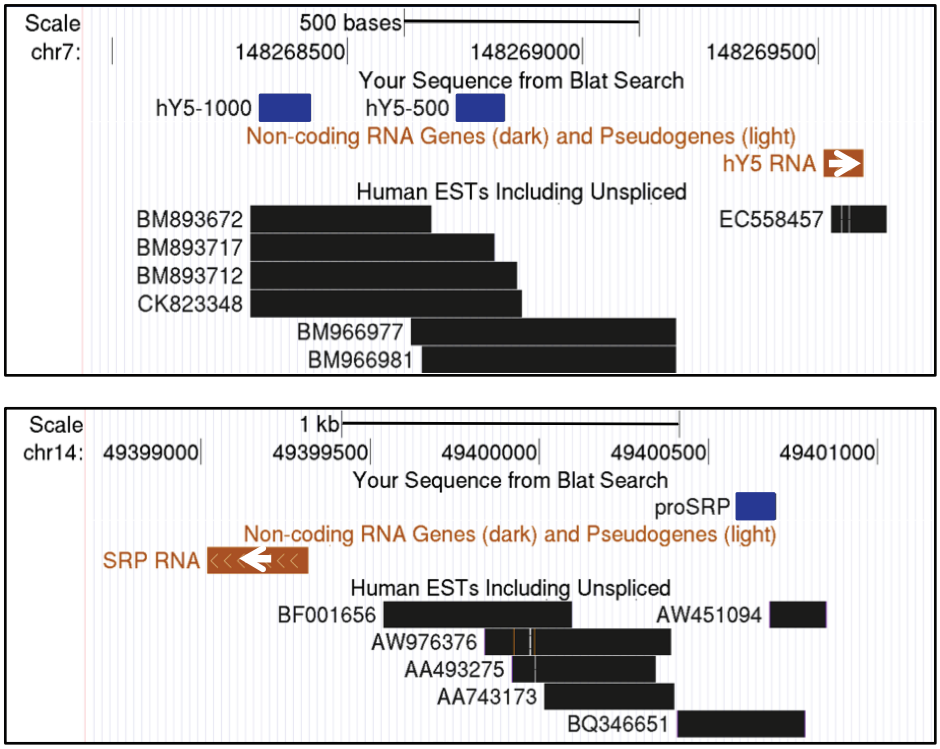


Preker et al., Sup. Fig. 3

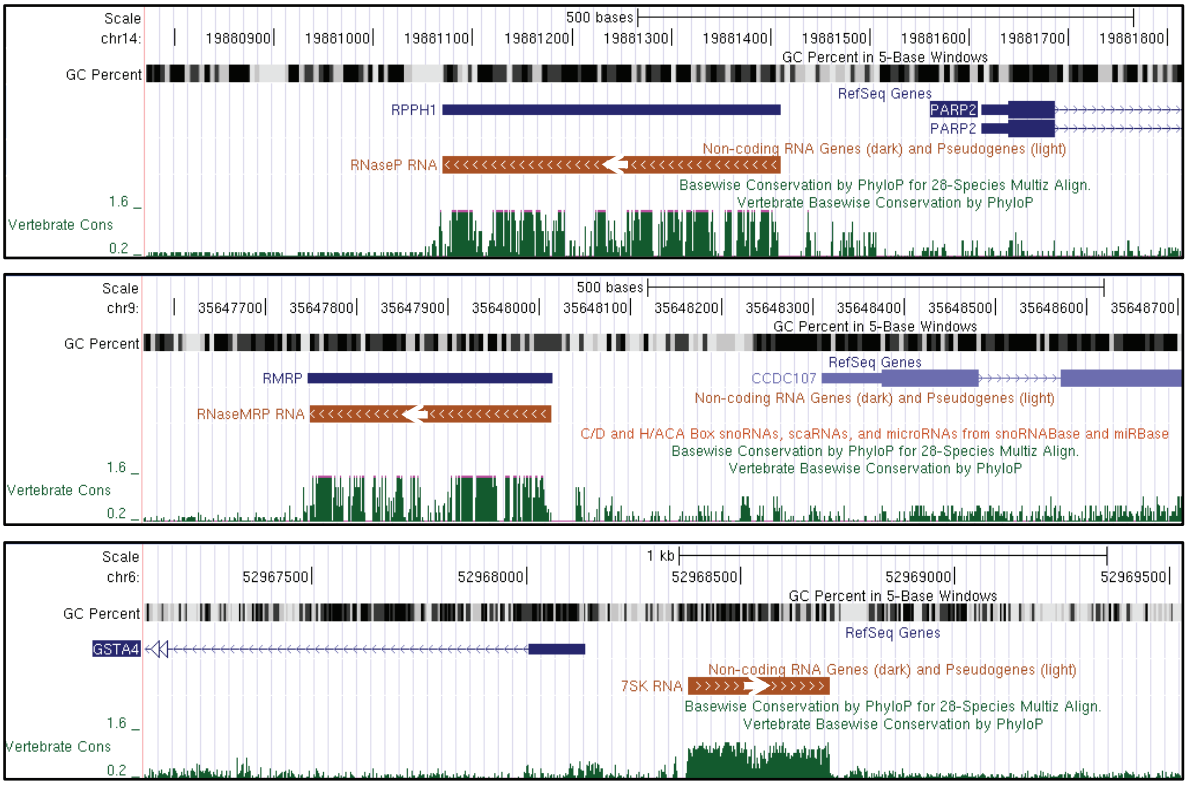


Preker et al., Sup. Fig. 4

A



B



Preker et al., Supplemental Table 1

target gene	sense siRNA	antisense siRNA
PAP-alpha	5'-gauuaggagugcauacaaatt-3'	5'-uuuguaugcacuccuaauctt-3'
PAP-gamma	5'-ggagaaacagaaaggaauatt-3'	5'-uauuccuuucuguuucucctt-3'
PAPD5	5'-gcgcugacguccagauauutt-3'	5'-aauaucuggacgucagcgctt-3'

Preker et al., Supplemental Table 2

Primer Pair	Orien-tation	Sequence#	Sources other than this Study	Alternative Name(s)
MYC				
MYC -2350	left	AAATCCAGTGTCTTGCTTTC		
	right	AATTCCAACAAACCCTAAAAC		
MYC -2140	left	CATCTCTTATGCGGTTGAAT		
	right	AGCCACCTCCTTGTTATTCT		
MYC -2000	left	GCTGGAAACTTGTTTTAAGG		
	right	TACTGGCAGCAGAGATCAT		
MYC -1760	left	TAAAGCTGAATTGTGCAGTG		
	right	GCTGTTCAAGATGGGTTATT		
MYC -1500	left	ATCATTTCAGGGAGCAAAC		
	right	ACTGTATGTAACCCGCAAAC		
MYC -1100	left	GTCCACAAGCTCTCCACTT		
	right	GTGCTAGACGGGAGAATATG		
MYC -600	left	CAAATGCAATGGGAGTTTAT	Marianne S.	msc33
	right	GTTTTCCTCCTTATGCCTCT	Christensen	msc34
MYC -500	left	ATCATTCTAGGCATCGTTTT		
	right	TGGAGGAAAGAAGGGTATTA		
MYC -250	left	GGAGCAGCAGAGAAAGGGAGA	Glover-Cutter et al,	MYC -400
	right	TGAATTAACTACGCGCGCCTA	2008	
MYC -200	left	GTAGTTAATTCATGCGGCTCTCTTACT	Glover-Cutter et al,	
	right	GGGCAGCCGAGCACTCTA	2008	
MYC +150	left	GGAGGGATCGCGCTGAGTA	Glover-Cutter et al,	MYC +1
	right	TCTGCCTCTCGCTGGAATTAC	2008	
MYC +300	left	AGATCCGGAGCGAATAG		
	right	CTGCTATGGGCAAAGTTT		
MYC +1400	left	GTTCCAGAACAGCTGCTAC		
	right	ACTCAATACGGAGATGCAA		
MYC +2220	left	TGCCCCTCAACGTTAGCTTC	Glover-Cutter et al,	MYC 2018
	right	GGCTGCACCGAGTCGTAGTC	2008	
MYC +2300	left	TGCAGCCGTATTTCTACT	Marianne S.	msc31
	right	GCTCGAATTTCTTCCAGATA	Christensen	msc32
MYC +5300	left	AAGTACATTTTGCTTTTTAAAGTTGATT	Glover-Cutter et al,	MYC 5155
	right	GGCTCAATGATATATTTGCCAGTTATTTTA	2008	
MYC +6200	left	TCCCCATATTAGAAGTAGAGAGGGAA	Glover-Cutter et al,	MYC 6028
	right	CTTGGGCATGTGGATGAGTCT	2008	
CCND1				
CCND1 -2500	left	CGTGAAGGTGATTTCAGTTA		
	right	TCTGTTCTTATTTCTTGGTGAC		
CCND1 -2300	left	TAAAACTGCAACCGAAAAGT		
	right	GCAGAGAAAGCATTTTGATT		
CCND1 -2000	left	CCTGGACGGCTCTTTAC		
	right	GCAGATCTCGACTAGGAACT		
CCND1 -1300	left	GATGACACTGAACTATATTC		
	right	TTACAGGTTCTGTCTCTTTG		
CCND1 -1000	left	GTCCCAGGGCAAATTCTA		
	right	TATTAGTCCCCTCTGCCTGG		
CCND1 -650	left	TTCGGGCATTTATTTTATTT		
	right	GGGGGTAACCCTAAAAGTT		
CCND1 -500	left	CTTGGGCATTTGCAAC		
	right	CACGTGTAAATTGCAAGAAC		
CCND1 -160	left	GCGATTTGCATTTCTATGA		
	right	CAAAAGCCATCCCTGAG		
CCND1 +1	left	GGCTTTGATCTTTGCTTAAC		
	right	GGAGGCTCCAGGACTTT		
CCND1 +300	left	GATGCCAACCTCCTCAAC		
	right	ACCTCCTTCTGCACACATT		

CCND1 +2000	left	GCTGGCCATGAACTACC		
	right	GGATGGTCTCCTTCATCTTA		
CCND1 +6950	left	TGTGAAGTTCATTTCCAATC		
	right	GTGTGAGGCGGTAGTAGG		

GAPDH				
GAPDH +1	left	CAAGACCTTGGGCTGGGACTG		
	right	GATGCGGCTGACTGTCGAAC		
GAPDH +55	left	CTCCTGTTCGACAGTCAGC	Glover-Cutter et al,	
	right	TTCAGGCCGTCCCTAGC	2008	
GAPDH +4511	left	AGATGTGTCAGGGTGACTTAT	Glover-Cutter et al,	
	right	TAGGTCCCAGCTACACGC	2008	
GAPDH-IVS	left	AGCTGCTGACCTTTCTGTAG		
	right	TTTTCCAGAAATCAGGAGTG		

Other RNAPII genes				
IFRG/RPT4	left	GCTCTTTATCTCTCTCTCAGCAAG		
	right	CAGTTTCGGTGTTCGGTTCA		
ID1	left	AGGTAAACGTGCTGCTCTAC	Marianne S.	msc35
	right	TGAAGGTCCCTGATGTAGTC	Christensen	msc36
NOP10	left	TCTCCCCAGATGACAAATAC		
	right	CTTGAAGCGTTTCTTGATG		
SMOC-IVS	left	CTGGGTGACAGAGTGAAACT		
	right	CCACCCTATTTTTGAGACAC		
β-glo +1	left	ATCGTCGACGAGCTCGTTTAG		
	right	CTAGTGAACACAGTTGTGTCAG		
β-glo mRNA	left	ACGTGGATGAAGTTGGTGGTG		
	right	TCTGATAGGCAGCCTGCACTG		

RNAPII PROMPTs				
40-2	left	GGGAGTCTAAGGAAAAGGAG	Preker et al, 2008	ProSTK11IP
	right	CAGTGAAAGGAGAGCGTATC		
40-9	left	GGCATCTGGACTAGAATGAA	Preker et al, 2008	ProSLC22A5
	right	TTGACACCGCCTAATCTTAT		
40-13	left	GGAAATAGTGGAGAAAAGCA	Preker et al, 2008	ProRBM39
	right	CATTTTTGAAGGAACGGTAG		
40-33	left	CTGGCCTAGCTAAAGTCTCA	Preker et al, 2008	ProEXT1
	right	TCTGCTCCTAGCTCTCAGTC		
40-52	left	AGTTCCAAGAAACCACACAC	Preker et al, 2008	ProPOGZ
	right	GGTCGTTTGAGTGGACTAAC		
40-53	left	AGGGTCAGAGTCCAAGTACA	Preker et al, 2008	ProCCDC93
	right	CTGGGTCGTATATGGAGAAG		
40-54	left	AAGGCCCCTACTTAACTCTC	Preker et al, 2008	ProIRF1
	right	GAGTTTTGGATGGAAAATGA		
40-56	left	TACCCTTCGGATACACTGTC	Preker et al, 2008	ProLACE1
	right	GGTACGCACTCAAATGTTTT		
40-57	left	AGATCCCAGTCATCTGTCAA	Preker et al, 2008	ProZNF800
	right	ATGTCACCAAGAGCAAGAAT		
40-58	left	TACCCAATTGGCCTAGTAAA	Preker et al, 2008	ProHMGN1
	right	CTGCCACTTTAAAGGTCTTG		
40-61	left	GGAGATACAGACCACGTTTT	Preker et al, 2008	ProSTAG2
	right	GAGAGCGAGCTGATTCTCTA		
40-64	left	GTCTTGGTCCTCCCTTTACT		ProIFNAR1
	right	CGATCCTGTGAAGGTCAA		
proFOS	left	CTCAGCGCAGATTTGAGT		
	right	GTCAGTTCGGGATGACAA		

RNAPIII genes				
U6	left	TTCGGCAGCACATATACTAA		
	right	GCCATGCTAATCTTCTCTGT		
hY5	left	CGAGTGTTGTGGGTTATTG		
	right	GGGGAGACAATGTTAAATCA		

SRP_N	left	GCCTGTAGTCCCAGCTACT		
	right	ACCATATTGATGCCGAACT		
SRP_C	left	GTGAATAGCCACTGCACTC		
	right	GTCTCGCTATGTTGCTCAG		
U6atac (chr9)	left	CAACACGCATACGGTTAAG		
	right	TTAGATGCCACGAAGTAGGT		
F-tRNA	left	GGAGAGCGTTAGACTGAAGA		
	right	ATTGAACCAGGGACCTTTA		
K-tRNA	left	GGATAGCTCAGTCGGTAGAG		
	right	CTGGACCCTCAGATTAAAAGT		

RNAPIII and RNAPI PROMPTs				
proU6 -630 (chr15)	left	TCCACCACTACTGGAAGTTT		
	right	TTCTCGGAACTAGCTCTCTG		
proU6 -1140 (chr15)	left	ATCAATGACAGCTCCAATGT		
	right	TGTGAAAGTGTGGCATAGAA		
prohY5 -500 (chr7)	left	TTTAAACTGCCGAAGAACTC		
	right	CCCCATAAGTCTACCCTCTT		
prohY5 -1000 (chr7)	left	TGCCTAAGTGCAGAAAATCT		
	right	TTTGGAATATCAGCCTCATC		
proSRP (chr14)	left	AAGAACCAGGGAGAAGAAAC		
	right	CTCCTAATTGGCAGGTCTAA		
proU6atac (chr9)	left	AAGGACCTGCCTTAACCTAC		
	right	TTGTGTGTGTTTCCAGTTGT		
proE-tRNA-CTC (chr18)	left	TAGGCCAACGCTGTTATAGT		
	right	CTTCTTGAGTTGGGAAACTG		
proK-tRNA-TTT (chr6)	left	TCCCCACCAGATTAGTTAGA		
	right	GCACCAATCGACACTGTAT		
proV-tRNA-TAC (chrX)	left	AAGATGGGCATAACATTGAG		
	right	TCACCTCTGGTCTCAGTCTC		
proHVG-1 (chr5)	left	ACGCCTGTAATCCCAATAC		
	right	GGATACTCTCGGTCTCTTGA		
pro-rDNA	left	GTTTTGGGCACCGTTTGTG	Sinkkonen et al, 2010*	42 kb
	right	GCGAAACCGTGAGTCGAGAA		

***S. cerevisiae* TOM1**				
TOM1 -930	left	TCCAAGGATGTTCTTCTGAG		
	right	GCTTCATAGGGTCTGGTATG		
TOM1 -1470	left	GCGTTGGTTGATTACTCATT		
	right	CTCAGAAGCAGCCATAAGTT		
TOM1 -1920	left	TTGGTTTCTTGGCTACTGTT		
	right	GGCTTTTATAATCGCTTCAAT		
TOM1 -2130	left	TCTTCGCCAACTACTTGATT		
	right	CGTGATAACTGACGGAAGA		
TOM1 -2700	left	ATGATGAAGAAATGGCAGAC		
	right	AGCGTCATCTTCATTTTCAC		
TOM1 -3030	left	GCAAGGTGAGAGTGAAGAAG		
	right	AATGGGTACGATTGATTGAG		

Oligonucleotides to confirm KDs				
PAP-alpha	left	CAACAGAACTCCACGTACAA		
	right	GACCACTCTGCCTTACTCAG		
PAP-gamma	left	ATCCGTGTCATCAAAAATTC		
	right	AAACAAGTGGCTGCATCTAT		
PAPD5	left	CAGCTGACCTCATCAAAGAT		
	right	GAACCAATTCCACCTGTAAA		

Oligonucleotide sequences used for quantitative PCR are shown in the 5' to 3' direction.

* Sinkkonen, L., Hugenschmidt, T., Filipowicz, W. and Svoboda, P. (2010) Dicer is associated with ribosomal DNA chromatin in mammalian cells. *PLoS One*, **8**, e12175.